8-2013

Correction of Cystic Fibrosis-specific Induced Pluripotent Stem Cells

Jacquelin Hanh Bui

Correction of Cystic Fibrosis-specific Induced Pluripotent Stem Cells

by

Jacquelin Hanh Bui, B.S.

Approved:

_____ **Supervisory Chair**
Brian R. Davis, Ph.D.

_____
Gilbert Cote, Ph.D.

_____
Naoki Nakayama, Ph.D

_____
Rick Wetsel, Ph.D

_____
Philip Ng, Ph.D

Approved:

_____
Dean, The University of Texas
Graduate School for Biomedical Sciences at Houston

Correction of Cystic Fibrosis-specific Induced Pluripotent Stem Cells

A DISSERTATION

Presented to the Faculty of

The University of Texas Health Science Center at Houston

and

The University of Texas M. D. Anderson Cancer Center

Graduate School of Biomedical Sciences

In Partial Fulfillment of the requirements for

The degree of

DOCTOR OF PHILOSOPHY

By

Jacquelin Hanh Bui, B.S.

Houston, Texas,

August, 2013

Acknowledgements

I am incredibly grateful to Dr. Brian Davis, my mentor, for taking a chance with a green, fresh-out-of-undergraduate, wannabe scientist. His enthusiasm for scientific inquiry is to be envied, if not emulated. However, I would be remiss if I did not acknowledge that he has taught me so much more than science; I know now, firsthand, what a good mentor/boss/leader looks like and can only hope to have another like him in my lifetime. There were many moments when Science was not kind to me. In those times, Dr. Davis both challenged and encouraged me, reminding me of my strengths while making me face my weaknesses. I do not think it an exaggeration to say that Dr. Davis' support kept me from quitting when everything else screamed that I was ill-fitted for a life of research. This degree is as much his as it is mine – not that he needs it. I am also thankful to my colleagues, Drs. Ana Maria Crane and Philipp Kramer, whose work and dedication to this project truly are responsible for its success, and to Wei Liao, who has supported our lab in ways immeasurable.

I would like to also express many thanks to my committee members: Drs. Rick Wetsel, Nayoki Nakayama, Phil Ng, and Gil Cote. They have been incredibly supportive throughout my journey. I could not have asked for a better group of scientists to sit at my biannual meetings; they have been a wellspring of direction and suggestions, all the while keeping in mind that I am but one person.

I owe so much to my family. My mom and dad, Le and Lang Bui, have sacrificed so much of their lives (and continue to do so) to give me the luxury of choosing an occupation, amongst other things; my sister and brothers (Holly, Victor, and Tony) have afforded me much laughter and reprieve from life's stresses; Jessica Ford, Hoang Dang, Vansa Shewakramani, and Scotty and Bethany Peckler have been so much more than just

friends, always enthusiastic and confident of my endeavors (even when I was not myself); and Alexander Griffith has quickly become my home these last two (and arguably, most uncertain) years of my graduate career.

I have many people to acknowledge for bringing me this far – many more than I have time or space to name. I have faith that each of these amazing individuals was brought into my life to share with me his/her own unique life, love, and laughter; for that, I am eternally grateful to our Creator.

Correction of Cystic Fibrosis-specific Induced Pluripotent Stem Cells

Publication No. _____

Jacquelin Hanh Bui, B.S.

Supervisory Professor: Brian R. Davis, Ph.D

Cystic Fibrosis (CF), affecting 1 in 3,500 live births in the US, is a disease caused by aberrant expression of the Cystic Fibrosis Transmemebrane Conductance Regulator (CFTR). While a multi-organ disease, CF-related complications and degradation of the lung is the leading cause of mortality. There are treatments to combat treatments but none that are curative. Recent advances in reprogramming and the ability to genetically modify the resulting induced pluripotent stem cells (iPSCs) have provided an alternative to conventional gene therapies thus far attempted for CF. Here we describe the *in vitro* generation of iPSCs starting from fibroblast cells obtained from a CF patient. These cells, like embryonic stem cells, show evidence of the ability to differentiate into any cell type. Then, using zinc finger nucleases (ZFNs) to induce a double strand break at the CF loci, we facilitated homologous directed repair of the break by providing a donor molecule that would, upon recombination, correct the disease-conferring mutation. In doing so, we observed an unpredicted allele-specific recombination, which could be applicable for gene correction of other diseases like those caused by dominant negative mutations. We interrogated the original CF fibroblast cells, the CF-iPSCs, and corrected iPSCs by whole genome and exome sequencing to evaluate genomic integrity after reprogramming and correction. Furthermore, *in vitro* differentiation of corrected iPSCs towards lung and thyroid lineages produces cells that express the mature CFTR protein isoform. This is the first report of site-specific correction of the *CFTR* gene in CF-specific iPSCs.

Table of Contents

List of Figures

List of Tables

Abbreviations

AAVS1	Adeno-associated virus integration site 1
ACTB	Actin, beta
ASL	Air surface liquid
B2M	Beta-2-microglobulin
bFGF	Basic fibroblast growth factor
BMP4	Bone morphogenetic protein 4
c-AMP	Cyclic adenosine monophosphate
c-MYC	v-Myc avian myelocytomatosis viral oncogene homolog
CCR5	C-C chemokine receptor 5
CF	Cystic Fibrosis
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CGD	Chronic granulomatous disease
Cl^-	Chloride
CMV	Cytomegalovirus
CNVs	Copy number variations
Cre	Cre-recombinase
CRIPTO	Teratocarcinoma-derived growth factor 1
DMEM	Dulbecco's modified eagle medium
dPBS	Dulbecco's phosphate-buffered saline
DSB	Double-strand break
EGTA	Ethylene glycol tetraacetic acid
ENaC	Epithelial sodium channel
ER	Endoplasmic reticulum

ESCs	Embryonic stem cells
FGF5	Fibroblast growth factor 5
FIAU	Fialuridine
FM	Fibroblast media
FOXA2	Forkhead box A2
FOXJ1	Foxhead box protein J1
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
HDR	Homology directed repair
HIV	Human immunodeficiency virus
HPRT1	Hypoxanthine phosphoribosyltransferase 1
IMDM	Iscove's modified Dulbecco's media
iMEFs	Irradiated murine embryonic fibroblasts
iPSCs	Induced pluripotent stem cells
KLF4	Kruppel-like factor 4
NHEJ	Non-homologous end joining
NKX2.1	NKK2 homeobox 1
OCT4	Octamer-binding transcription factor 4
PAX8	Paired box 8
REX1	RNA exonuclease 1 homolog
RP13A	Ribosomal protein L13a
SFD	Serum-free differentiation meda
SNP	Single nucleotide polymorphism
SOX17	Sex determining region Y-box 17

SOX2	Sex determining region Y-box 2
TALENs	Transcription-like effector nucleases
UTR	Untranslated region
VSV-G	Vesicular stromatitis virus
WNT3a	Wingless-type MMTV integration site family, member 3a
wt	Wild-type
X-SCID	X-linked severe combined immunodeficiency
ZFN-L	Zinc-finger nuclease left
ZFN-R	Zinc-finger nuclease right
ZFNs	Zinc-finger nucleases
ZFPs	Zinc-finger proteins

Chapter 1: Introduction

1.1 Cystic Fibrosis

Cystic Fibrosis (CF) is the most common autosomal recessive disease among Caucasian populations, with an incidence of about 1 in 3,500 (*1*) CF is a monogenic disorder with CF carriers showing no clinical symptomology. The CF gene is located on chromosome 7 and encodes a 1480 amino acid protein, Cystic Fibrosis Transmembrane Conductance Regulator (CFTR). CF is caused by any of more than 1000 mutations (*2*) in the *CFTR* gene that fall into 6 classes resulting in either: I) no CFTR synthesis, II) defective processing, III) misregulation, IV) altered conductance, V) partly defective production or processing, or VI) defective regulation of other channels (*3*). The most prominent mutation, ΔF508, is a class II mutation. The ΔF508 mutation is a deletion of three nucleotides resulting in a deletion of phenylalanine at amino acid position 508 and occurs in ~70% CF alleles (*4*). Another severe mutation is ΔI507, which is prominent in a population in north west France (*5*); compound heterozygotes with genotype ΔI507/ΔF508 have also been documented (*6*, *7*).

As a member of the ATP binding cassette transporter family, the CFTR protein has a regulatory domain, two nucleotide-binding domains, and two membrane-spanning domains (*8*, *9*). Upon reaching conformational maturation during normal protein production, the wild-type (wt) CFTR protein is coated with COPII in the endoplasmic reticulum (ER), which leads to budding from the ER and transport to the Golgi for further processing (*3*). The ΔF508 CFTR protein, although synthesized, fails to fold correctly and does not achieve the mature, glycosylated form. Maturation of the wild-type protein has proven to be inefficient; even more so for the ΔF508 protein. Approximately 75% of

wt protein and 100% of ΔF508 protein do not reach structural maturity and are degraded in the ER by the ubiquination-proteasome system (*10*).

1.1.1 CF pathophysiology

CFTR is normally expressed on the apical surface of various epithelial tissues, including the lung, sweat glands, intestine, pancreas, and reproductive organs (*4*, *9*, *11-13*). Though a complex disease affecting many organs, CF-associated lung disease is the major cause of morbidity and mortality (*14*). In the lung, CFTR functions as a cAMP-activated chloride channel. The prevailing hypothesis of how the ΔF508 mutation results in the pathophysiology of CF is that the absence of mature CFTR protein results in aberrant chloride transport and epithelial sodium channel (ENaC) upregulation, causing increased sodium and water reabsorption (*2*, *4*, *15*). Resulting dehydration of the air surface liquid (ASL) and thickening of the mucous covering the lung epithelium leads to impaired mucous clearance, thereby creating a microenvironment in which bacteria thrive. Patients thus suffer from recurring bacterial infections and chronic inflammation, which eventually leads to respiratory failure and death. Other CF symptoms include malabsorption in the intestine, pancreatic insufficiency, and increased chloride and sodium levels in the sweat (which is used often to diagnose CF). Furthermore, 91-98% of CF males are sterile due to defective or absent vasa deferentia (*16*).

1.1.2 Treatments

Current treatments for CF act primarily to ameliorate and manage symptoms of the disease. CF patients, for example, maintain a strict and rigorous regimen of nutrition to overcome malnutrition resulting from abdominal obstruction; chest physiotherapy to loosen the mucous and facilitate clearance; antibiotic treatment upon bacterial infection;

and inhaled steroids or non-steroidal drugs to fight inflammation (*2*, *3*, *17*). As a result of improvements in these treatments, the median survival rate of CF patients has increased to 35 and 40 years; by comparison, in the 1970s CF patients were not expected to survive beyond adolescence (*18*). Nonetheless, the quality of life for CF patients remains dampened due to the rise of drug-resistant bacteria and disease related complications in other organs (such as liver disease, diabetes, and fertility issues) (*19*). Comprehensive treatment of CF would have to address the disease in these other tissues, as well as provide more curative (rather than symptomatic) approaches for lung disease.

To get at the underlying issue (i.e. aberrant CFTR protein expression at the cell surface), there are a number of studies focused on chemical chaperones and small molecules that rescue ΔF508 CFTR trafficking, such as glycerol, genestin and benzoquinolizinium compounds (*10*, *20-22*). However, it has been shown that, though the ΔF508 CFTR protein can be facilitated to reach the apical surface, the mutant protein has a decreased half-life (*10*). Furthermore, evidence that processing of ΔF508 CFTR is tissue dependent, with differences even within the lung, suggests that bypassing ER quality control and trafficking alone may be insufficient for treatment of CF (*23*).

Since the CFTR gene was discovered in 1989 (*4*), gene therapy has been considered as a potential therapeutic approach to curing CF. There have been attempts to deliver the CF gene by lentivirus (*24*, *25*), adenovirus (*26-28*) and adeno-associated virus (*29*, *30*). However, several barriers – both immune and mechanical – have prevented these attempts from being clinically beneficial. In the normal lung airway, for example, natural defenses against pathogens (i.e. mucous and mucociliary clearance) also serve as effective barriers to delivery of gene therapy reagents. This is exacerbated in a CF lung

where decreased ASL and thickened mucous create a barrier that is more impervious. Also, chronic bacterial infection leads to increased numbers of neutrophils, which also serve as a hindrance to viral delivery of gene therapies (*27*, *31*). Another complication is that most viral receptors are localized basolaterally (*32*, *33*). Though this can be circumvented by using agents like EGTA to transiently disrupt tight-junctions and, thereby, make the receptors accessible (*28*, *34*), doing so could worsen infection and increase the risk of systemic bacterial infection (*1*). Furthermore, in clinical trials in which delivery the CF gene to patients was accomplished by either adenovirus (*26*) or adeno-associated virus (*35*), there has been evidence of cellular and humoral adaptive immune responses against repeated administration with no increase in gene transfer.

In addition to overcoming these barriers, a critical consideration is the degree to which CFTR expression needs to be restored (i.e. percentage of cells expressing) for successful gene therapy of CF. In initial studies of mixed populations of a CF airway cell line and retrovirally-corrected airway cell line, epithelial cultures consisting of 6-10% of corrected cells was sufficient to restore chloride transport to levels of 100% of corrected cells (*36*). However, to restore ENaC activity to normal levels, more than 90% of cells required CFTR overexpression (*37*). One drawback of these studies is that they were performed in airway cell lines, which may not replicate *in vivo* epithelium. To overcome this hurdle, Farmen et al. used mixtures of freshly isolated wild-type and CF airway cells at varying proportion and found that a 20% wild-type and 80% CF mixture generated 70% of the Cl^- current that 100% wild-type epithelium did, indicating that even a 20% correction would be clinical beneficial (*10*). While the percent of cells that need to be corrected are still low, there is some concern that “supercorrection” of CFTR in these

cells may lead to mislocation of CFTR to the basolateral membrane, though this too has been refuted (*10*, *38*).

1.1.3 Cellular expression of CFTR

Because CF morbidity and mortality is primarily associated with lung failure, characterization of CFTR expression within the lung has been explored extensively. The respiratory system is a complex branching organ consisting of proximal and distal parts. The proximal part of the lung includes all the major conducting airways: the nose, trachea, and bronchi. These structures are lined with a psuedostratified columnar epithelium consisting of ciliated and non-ciliated cells, basal cells, goblet cells and submucosal glands (FIGURE 1.1). The distal airway includes the bronchioles and alveoli, in which gas exchange occurs. A simple ciliated epithelium with Clara cells, submucosal glands, and a few goblet cells line the bronchioles. As one moves distally, the ratio of Clara cells to ciliated cells increases. Finally, the epithelial lining the alveoli consists only of two cells types, alveolar type I and II cells.

Unfortunately, characterization of CFTR in the respiratory cells/tissue has resulted in contradictory findings. In early studies, Engelhardt and collegues made compelling arguments for the submucosal glands as the predominant site of CFTR expression (*39*). They also detected *CFTR* mRNA by *in situ* hybridization and protein by immunofluorescence in non-ciliated cells of the bronchioles and alveoli (*40*). However, several more recent studies have localized CFTR expression to the ciliated cells, not the submucosal glands (*13*, *41*). These contradictions are probably due to experimental complications, including issues with antibody sensitivity as some antibodies have inaccessible epitopes after routine tissue preservation. Most recently, Kreda et al.

FIGURE 1.1 Lung structure and cell populations. Though there are cellular differences between murine and human airways, most studies on adult lung, particularly those attempting to identify lung stem cells, have used mouse models. The human lung consists of proximal (trachea and bronchi) and distal (bronchioles and alveoli) branches of differing epithelial organization to suit function. The proximal trachea and bronchi are composed of a pseudo-stratified epithelium of ciliated and non-ciliated cells. There is evidence for two populations of progenitor stem cells (denoted with an asterisk) in the trachea and bronchi: basal cells and serous cells within the submucosal glands. In the bronchioles, the basal cells are replaced by Clara cells, which are the putative progenitor cell population for this region. The epithelium in the alveoli is much simpler with just alveolar type I and type II cells, the latter believed to have progenitor capacities. In addition to decreased complexity of epithelial organization as one moves distally towards the alveoli, CFTR expression is evidenced to decrease as well.

CFTR
Non-ciliated
Ciliated
Goblet
Basal*
Sub-mucosal gland*
Cara*
Ciliated
Secretatory gland*
Type II
Type I*
Trachea
Bronchi
Bronchioles
Alveoli

reported *in situ* hybridization of *CFTR* mRNA and immunolocalization of CFTR protein in nasal and bronchial airways was greater in the superficial epithelium than the submucosal glands (*42*). Here, laser confocal microscopy was used to detect immunostaining by anti-CFTR antibodies reported to have 10-100 higher sensitivity than those used previously (*43*).

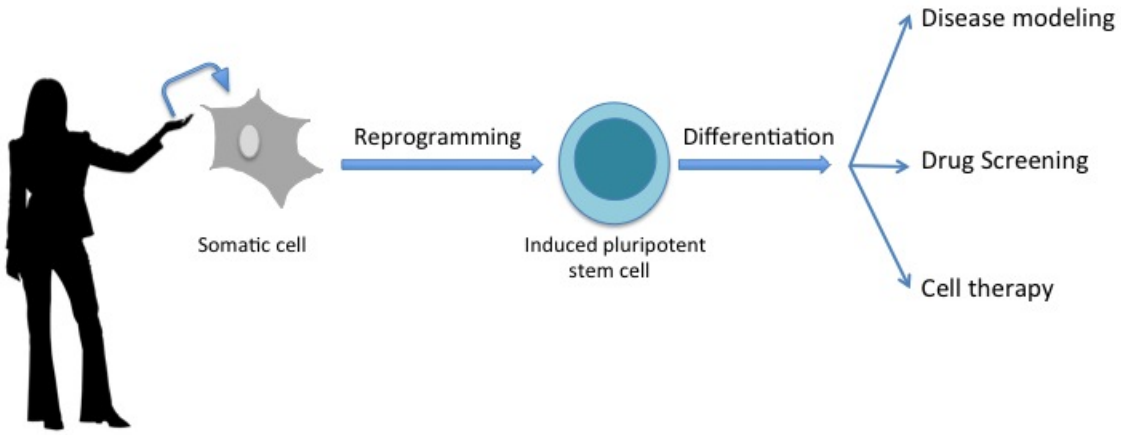
1.2 Induced Pluripotent Stem Cells

Induction of pluripotency (the ability to differentiate into all three germ layers) in somatic cells by defined factors *in vitro* holds great promise for the field of regenerative medicine as well as for disease modeling and drug testing (FIGURE 1.2). The appeal of these induced pluripotent stem cells (iPSCs) for regenerative medicine is 1) the potential for transplantation of not only terminally differentiated cells but also stem progenitors and 2) that these transplanted cells are autologous, having been derived from the patient's own cells. iPSCs may provide both short- and long-term cellular therapy with little to no risk of immune rejection based on histo-incompatibility.

1.2.1 Generation of iPSCs

When reprogramming was first reported in 2006 by Takahashi et al., pluripotency was induced in murine fibroblast cells by ectopic expression of four transcription factors: *Oct4*, *Sox2*, *Klf4*, and *c-Myc* (*44*). By 2007, two independent groups quickly reproduced this by reprogramming human fibroblast cells (*45*, *46*) and, since then, iPSCs have been generated from several tissues including keratinocytes (*47*), pancreatic beta cells (*48*), and B lymphocytes (*49*). Furthermore, a myriad of disease-specific iPSCs have been generated, including Cystic Fibrosis (*50-53*).

FIGURE 1.2 Potential applications for induced pluripotent stem cells (iPSCs). Somatic cells like skin fibroblasts, which can be easily obtained, can be reprogrammed to induce an embryonic-like pluripotent state by ectopic expression of a defined set of factors. These iPSCs can be used for disease modeling either by reprogramming disease-specific cells or by introducing disease conferring mutations; followed by differentiation to regenerate somatic cells used for drug screening and toxicity testing; and autologous (patient-specific) cellular therapy.



Ectopic expression of the reprogramming factors was first achieved by retroviral transduction. This presented several barriers for clinical use of such cells. For example, though one of the hallmarks of a fully reprogrammed cell is that the transgenes are silenced and the endogenous pluripotential network genes reactivated (*54-56*), Okita et al. found that 20% of the adult chimeras generated from iPSCs formed tumors due to reactivation of *c-myc*, a known oncogene (*57*). This has led to reprogramming without c-Myc; although such 3 factor reprogramming is of lower efficiency (*58*). While 3 factor reprogramming efficiency can be improved by treatment of somatic cells with valproic acid (*59*), integration of exogenous reprogramming factor cDNAs – *c-Myc* included or not – still raises concerns of insertional mutagenesis as it has been demonstrated that reprogrammed cells can have anywhere from 10 to 20 proviral integrations (*45*, *54*, *60*). Several groups have now demonstrated the ability to reprogram human cells using a variety of methods that reduce or eliminate the risk of insertional mutagenesis, including the use of excisable polycistronic lentiviral vectors (*61-63*), non-integrating Sendai virus (*64*), *piggyBac* transposon (*65*), nonintegrating episomal vectors (*66*), and recombinant proteins (*67*). Interestingly, Somers et al. demonstrated that, using an excisable polycistronic lentiviral system, they were able to reprogram cells with only one integration of the vector (*63*).

1.2.2 Relationship between iPS and embryonic stem cells

Like embryonic stem cells (ESCs), iPSCs express key stem cell markers and demonstrate the ability to form teratomas containing cell types of all three germ layers, evidence of a pluripotent state. In the case of mouse iPSCs, they can be used to derive chimeric mice (*49*, *58*). It has also been demonstrated that these iPSCs have global

patterns of histone methylation similar to ESCs and reprogramming of female somatic cells reactivates the inactive X-chromosome (*54*).

Though highly similar to ESCs, several reports have indicated that iPSCs may be distinguishable from ESCs. For example, using high resolution SNP to analyze primary human fibroblast line and various passages of the iPSC clones generated from that line, a higher number of copy number variations (CNVs) was found compared to non-pluripotent cells (*68*). Laurent et al. found the reprogramming process associated with deletions in tumor suppressor genes, some of which were selected against with long-term passaging. Hussein et al. likewise found that some CNVs arising *de novo* in early passage iPSCs have a selective disadvantage with further passaging, driving the lines towards a genetic state resembling that of ESCs (*69*). There is also evidence of selection for chromosomal aberrations in both ESCs and iPSCs over prolonged culture (*70-72*) and that iPSCs have additional abnormalities as a result of the reprogramming process itself (*73*).

At this juncture, it remains to be seen how similar ESCs and iPSCs truly are and whether comparison to ESCs is even the best marker for potential application of iPSCs. What is clear is that clinical application of human iPSCs will require improvements in generation and maintenance of iPSCs as well as frequent assessment of genomic stability.

1.2.3 Differentiation of iPSCs

Human iPSCs have now been used to generate a variety of cell types, for example insulin-producing pancreatic cells (*74*), adipocytes (*75*), and motor neurons (*76*). Recently, a group has reported generation and transplantation of human iPSC-derived retinal pigmented epithelium into mice and demonstrated improved visual function and

no tumor development over their lifetimes (*77*). Generation of lung cells from pluripotent cells have been approached in various ways. Wang et al., for instance, reported the derivation of alveolar type II cells from human ESCs that were stably transfected with an ATII-specific surfactant protein C promoter-neomycin transgene; the ESCs were spontaneously differentiated and selected for neomycin-resistant ATII cells (*78*).

Another approach to differentiation is one that attempts to recapitulate, in a step-wise manner with exogenous factors, the normal embryonic development of the desired cells. Using this approach, two independent groups reported differentiation of mouse ESCs and iPSCs into early multipotent $Nkx2.1^+$ cells (which developmentally give rise to lung and thyroid) and airway progenitor cells. Longmire et al. seeded decellularized mouse lung tissues with these mouse ESC-derived $Nkx2.1^+$ cells and showed engraftment of cells that lost Nkx2.1 expression but expressed alveolar type I protein T1α (*79*). Mou et al. also differentiated mouse ESCs into $Nkx2.1^+$ cells and showed that these cells were capable of giving rise to basal, Clara, ciliated, and goblet cells when subcutaneously engrafted into immune deficient mice; human ESC-derived $Nkx2.1^+$ cells only showed evidence of basal cell differentiation when likewise transplanted (*80*). Recently, Wong et al. reported the derivation of tissue from human iPSCs and ESCs that they claim to be mature airway (*81*). In this report, they noted the presence of tight junctions as well as the apical localization of CFTR by immunofluorescence. Furthermore, in a functional assay for CFTR, they evidenced that the generated cells include appropriately expressed and functional CFTR.

1.3 Genetic Correction of iPSCs

Gene therapy is commonly accomplished by viral delivery of a copy of the correct gene. Though gene therapies using randomly integrating virus to deliver the transgenes have been successfully applied in clinical trials for treating X-SCID (*82*) and Chronic Granulomatous Disease (CGD) (*83*), these promising results are overshadowed by some of the treated patients acquiring leukemia, in the case of the X-SCID treatments, or myelodysplasia, in the case of CGD. In both instances, these adverse effects resulted from random integration of the transgene adjacent to proto-oncogenes.

In principle, site-specific integration of the correct gene would circumvent complications arising from random integration. Gene editing by homologous recombination allows researchers to edit in a site-specific manner. The goal of such site-specific correction is to alter just a couple of bases of the 6×10^9 base pairs (bp) in a diploid human genome. Gene targeting and editing by homologous recombination can either be targeted to a safe harbor site (a constitutively expressed gene, disruption of which does not perturb normal function of the manipulated cell) or to the particular locus where the disease-conferring mutation is found. The latter is arguably favorable over the former because it does not generally require the full coding sequence nor regulatory sequences. Instead, in the latter case, only the mutations are corrected and expression of the gene should still be regulated by all the transcriptional regulatory elements of the endogenous gene.

However, homologous recombination, regardless of site, happens infrequently – typically 10^{-7} to 10^{-5} in mouse ESCs (*84*). Fortunately, a double-strand break (DSB) at the target site can increase site-specific gene targeting events up to 1000-10,000 fold (*85*, *86*). DSB-mediated gene editing can be achieved by homology directed repair (HDR),

provided that there is a template of sufficient homology to the region immediately adjacent the DSB. In the absence of a homologous donor, non-homologous end joining (NHEJ) can be utilized to introduce mutations. After pioneering work by Jasin and colleagues in utilization of meganucleases like I-*Sce*I to create site-specific DSBs for gene editing (*87*), researchers have been investigating other DNA nucleases that can be easily customized to any gene of interest. Currently, one of the most widely studied sequence-specific nucleases are zinc finger nucleases (ZFNs).

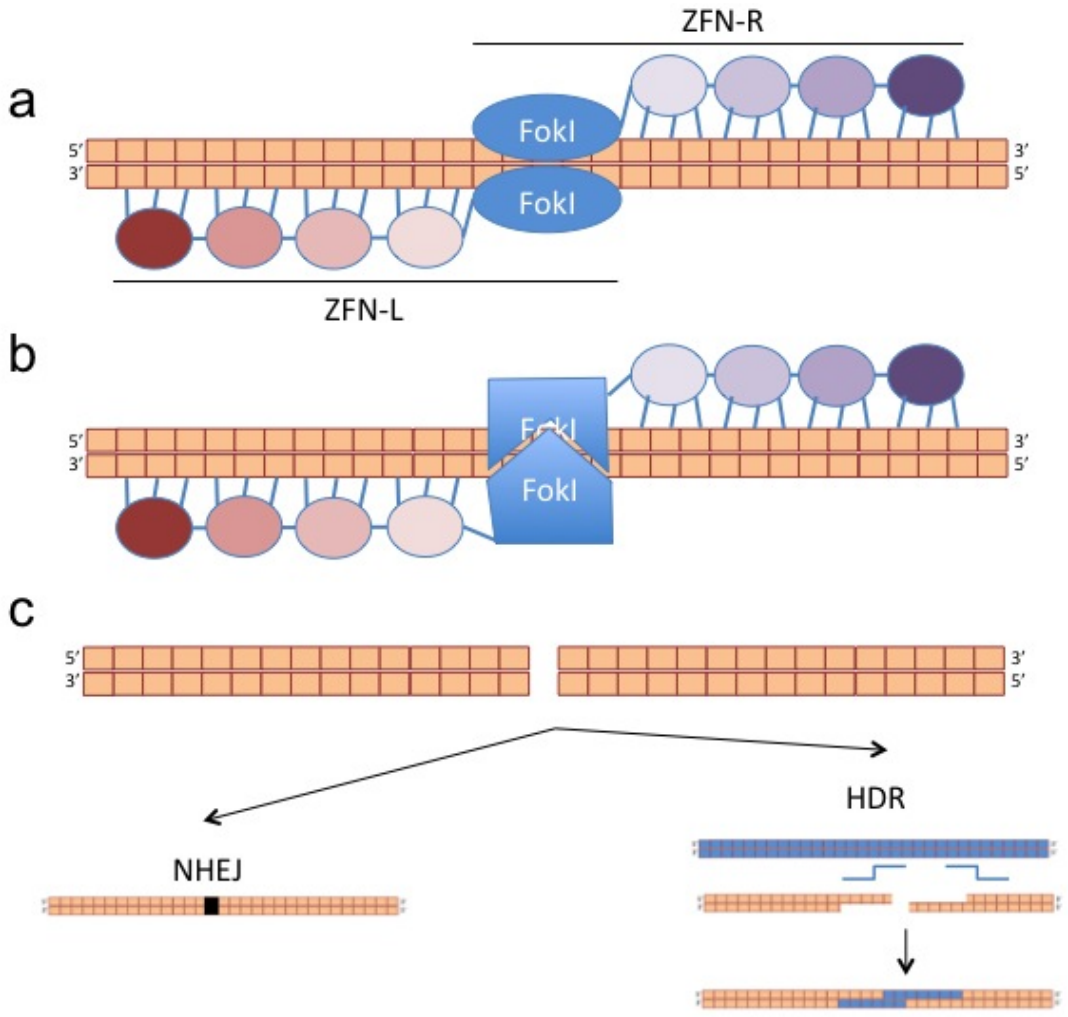
1.3.1 Zinc finger nucleases

A zinc finger nuclease is a fusion of zinc finger proteins to the cleavage domain of *FokI* (FIGURE 1.3a) (*88*). *FokI* is an endonuclease consisting of an N-terminal DNA-binding domain and a C-terminal domain with nonspecific cleavage activity. The zinc finger proteins (ZFPs) are a class of DNA-binding proteins, with each ZFP interacting with three nucleotides and acting as relatively independent modules (*89*). Initially, ZFNs were designed to bind a predetermined site of 9-12 bp by linking 3-4 zinc finger proteins together. Because *FokI* requires dimerization for cleavage, ZFNs work as a pair with each half-site (ZFN-L and ZFN-R) binding on opposite strands and separated by a space of 5 or 6 bp to allow for optimal dimerization of the *FokI* domains (*90*).

1.3.2 Specificity and optimization of ZFNs

Specificity of ZFNs, then, is conferred by two sets of specific sequences (for a pair of ZFN, each with 3 zinc finger proteins for example, this is two sets of ~9 bp for a total of ~18 bp) and the probability that these two sets are in close enough proximity to allow dimerization of the *FokI* domains. Still, staining ZFN-treated cells with antibodies against proteins, like 53BPI, that are recruited to DNA damaged foci for repair reveals

FIGURE 1.3 Zinc Finger Nucleases (ZFNs) can create site-specific double strand breaks to induce genome modifications. a) ZFNs are bioengineered by linking 3 or more zinc finger DNA binding domains (4 shown above), which each recognizes and bind 3 specific bases, with the sequence-*un*specific cleavage domain of *FokI*, which requires dimerization for activity. Functional only as a pair, ZFN specificity is then conferred by 9+ sequences on one side or "arm", 9+ sequences on the other arm, and the need for the two arms to be close enough in proximity to provide a spacer region in which the two *FokI* domains can dimerize and cut. b) Early configurations of ZFNs could become active by any dimerization of two ZFNs: ZFN-L/ZFN-L or ZFN-R/ZFN-R, as well as the intended ZFN-L/ZFN-R. Newer ZFNs have *FokI* domains that have been modified such that they act only as obligate heterodimers; only a ZFN-L/ZFN-R configuration is functional, thereby increasing specificity. **c)** Double strand breaks can be repaired by two means: non-homologous end joining (NHEJ), which is error prone (illustrated as the introduction of a mutation depicted in black), or homology directed repair (HDR), which requires a homologous template (shown in blue).


a
ZFN-R
FokI
FokI
5′
3′
3′
5′
ZFN-L
b
FokI
FokI
5′
3′
3′
5′
c
5′
3′
3′
5′
HDR
NHEJ

tens of foci when there should only be two (if there is biallelic cutting), thus implicating unintended ZFN effects (*91*, *92*). By allowing up to 2 mismatches per half-site in the targeted binding sequence for ZFNs targeting the *CCR5* gene, which encodes a major coreceptor for HIV-1 entry, Perez et al. found that the ZFNs also modified two other sites, one of which is the closely related *CCR2* gene (*93*). Furthermore, it has been demonstrated that there is ZFN-induced toxicity due to binding infidelity (and subsequent cleavage) (*94-96*).

For therapeutic purposes, then, it is pertinent to reduce ZFN off-target activity and toxicity. To address this issue, some groups have increased the number of zinc finger proteins linked to make one ZFN (*97*). Hockemeyer et al., for example, used ZFNs containing 4-6 zinc finger proteins for a combined recognition for the ZFN pair of 24-36 bp (*98*). In addition, although ZFNs are modular in design, the binding specificity of the individual ZFPs have been shown to be context-dependent, i.e. affected by adjacent ZFPs. Thus, increased DNA-binding specificity has been achieved by optimization of adjacent ZFP pairs, rather than as individual ZFPs (*99*).

Other factors may also play a role in defining specificity of a ZFN. For example, two groups have demonstrated that decreases in ZFN concentration also decrease off-target cutting. Gupta et al. found that putative off-target sites that were actually cut share common characteristics with the intended target sequence: conservation of guanines (*100*). In putative off-target sites that were not cut, these same positions were found more diverse. Pattanayak et al. found that suboptimal molecular interactions at particular target site positions in one half-site led to reduced tolerance of mutations in the other half site,

even at other positions within the same half-site (*101*). Together, these results suggest ways that zinc fingers can be engineered for higher specificity.

Alternatively, the ZFN structure can be altered at the cleavage domain to increase specificity. As mentioned before, the *FokI* nuclease domain requires dimerization to be active. ZFNs are created to function as a pair, each ZFN with its own unique DNA-binding domains; however, theoretically, homodimers of ZFN-L/ZFN-L or ZFN-R/ZFN-R can also bind and cleave. Three groups have reported alteration of the nucleases' dimerization interface to create variants that are functional only as obligate heterodimers (FIGURE 1.3b), thus abrogating the risk of cutting by homodimers while maintaining the advantages of cleavage that is dimerization-dependent (*91*, *92*, *102*).

1.3.3 ZFN-mediated homology directed repair

Donor-free delivery of ZFNs can be used to disrupt genes by inducing NHEJ (*84*, *95*). Alternatively, gene addition can be achieved by providing an appropriate donor template along with the ZFNs, taking advantage of HDR; such approach has been used to correct genetic diseases (FIGURE 1.3c). In fact, gene correction can be accomplished directly at the mutated gene (*103*) or at a safe-harbor locus like *AAVS1* (*104*, *105*). Furthermore, biallelic modifications can be made (*93*, *106*).

The gene targeting efficiency may depend on several factors including accessibility of the target gene and how the ZFNs and donor are delivered (*98*). In addition, the design of the donor definitely plays a role. Pruett-Miller et al. optimized their donor template by altering the sequences recognized by the ZFNs to prevent ZFN cutting of the newly repaired gene (*92*). Highly influential to the efficacy of the donor as a template for HDR is the extent of homology between the targeted locus and the

targeting (donor) vector. Most of the information relevant to length and extent of homology between donor and target locus come from standard HR in the absence of an introduced sequence-specific DSB. It seems that the total length of DNA sequence present in the donor is not as important as the amount of homology between it and the target sequence (*107*). A mere 2-fold increase in homology increases gene targeting frequency by 20-fold (*108*). Waldman and Liskay found that there was little difference in recombination frequency between sequences with 232 bp of uninterrupted homology and the frequency of recombination between two sequences with four mismatches distributed throughout while still maintaining a 232 bp stretch of homology (*109*). Similar experiments studying the effects of donors with various amounts of heterology on recombination frequency in the presence of an introduced DSB (via SceI) support these conclusions (*110*).

1.3.4 ZFN-mediated correction of iPSCs

Recently, there have been multiple reports of ZFN-mediated gene editing of iPSCs (extensively reviewed in (*111*)). Hockemeyer et al. have used ZFNs to target both silenced and expressed genes including the safe harbor *AAVS1* locus (*98*). Another locus, the *CCR5* locus, has been routinely used for stable transgene expression in a site-specific manner (*112*). ZFNs have also been used to create mutations. Soldner et al., for instance, used ZFNs to introduce two different point mutations associated with Parkinson's disease to create an *in vitro* disease model (*53*). Alternatively, ZFNs can be used in iPSCs to correct diseases like α1-antitrypsin deficiency (*113*), sickle cell disease (*114*), and α-thalassemia (*115*). For example, to correct α-thalassemia, Chang and Bouhassira accomplished gene correction by insertion of a goblin transgene into the *AAVS1* safe

harbor site of iPSCs and showed that these corrected iPSC were able to give rise to erythroid cells. Such reports suggest that ZFN-mediated gene edits in iPSCs do not alter differentiation potential.

1.3.5 Other tools for targeted genome editing

Gene targeting facilitated by ZFNs and homing nucleases, a meganuclease like I-SceI, have been well documented. Still, the search for more easily engineered, cheaper, and safer tools for enhancing gene targeting continues. Recently, transcription activator-like effector nucleases (TALENs) and RNA-guided Cas9 nucleases have been added to that list. Due to their flexibility in design, both TALENs and Cas9 are reported to be capable of targeting more of the genome than ZFNs and have now been used to genetically modify human iPSCs (*116-118*). Ultimately, a variety of factors will determine which nuclease platform will be used for a given therapeutic application – principally the efficiency of cleavage and absence of off-target effects.

1.3.6 Gene correction of CF

Several factors make CF a prime candidate for targeted genome correction: 1) CF is a monogenic disease, and 2) it follows recessive Mendelian inheritance with carriers showing no clinical manifestations of CF so correction of just one allele would restore carrier state. To ensure appropriate control from endogenous transcriptional regulatory elements, an endogenous approach (vs random integration or safe harbor) would be more applicable. One option for endogenous correction of CF is to knock-in *CFTR* cDNA to provide a full copy of the correct gene that would correct a broad range of CF mutations. However, 70% of CF alleles have a deletion of just three nucleotides, ΔF508 (*119*); thus correction of this particular mutation would still benefit a majority of CF patients.

Restoration of the three nucleotides missing in ΔF508 CFTR should restore proper protein folding, allowing for proper glycosylation, maturation, and thus transport to the Golgi and, then, to the cell surface. Correction of CFTR folding and maturation can be detected by western blot. Fully mature CFTR protein is ~170 kDa and is sensitive to deglycosylation by *N*-glycosidases, whereas the ΔF508 mutation results in an immature core-glycosylated protein that is ~140 kDa.

1.4 Objectives and Hypothesis

We hypothesize that we can utilize these two novel technologies – reprogramming of somatic cells by defined factors and ZFN-mediated gene editing of iPSCs – to provide the foundation for an alternative to conventional gene therapy approaches for treating Cystic Fibrosis. We hypothesize that patient-specific iPSCs can be generated from fibroblast cells obtained from CF patients and corrected to restore wild-type CFTR expression in iPS-derived lung cells. To accomplish this, we aimed to 1) generate CF-specific iPSCs, 2) correct the disease-conferring mutation, and 3) demonstrate that genetic correction will lead to restored protein expression.

1.4.1 Specific Aims

To accomplish our goals, three specific aims were proposed:

SAi. Generate a line(s) of induced pluripotent stem cells (iPSCs) from primary fibroblasts derived from Cystic Fibrosis patients. Fibroblast cell lines derived from patients with CF were reprogrammed by ectopically expressing reprogramming factors, delivered by viral means. To assess similarity to embryonic stem cells, reprogrammed cells were characterized based on cell and colony morphology, expression of stem cell markers by immunohistochemistry, expression of stem cell genes via real-time PCR

assay, and the ability to form teratomas in immunodeficient mice injected with the human iPSCs.

SAii. Correct the disease-conferring genetic mutation in the iPSCs using zinc finger nuclease (ZFN) -mediated homologous recombination. We employed ZFNs targeting exon 10 of CFTR to induce homologous recombination of the endogenous, deleted sequences with a donor containing the wild-type sequences to correct the mutation in reprogrammed cells. To demonstrate successful gene correction in a targeted manner, PCR and Southern blot analysis were used to confirm site-specific integration and ensure no off-target insertion site was generated.

SAiii. Determine restored CFTR expression in lung cells differentiated from genetically corrected iPSCs. To demonstrate that our method of correcting *CFTR* results in restored CFTR expression, we compared the forms of CFTR that are generated from lung cells derived from uncorrected vs. corrected iPSCs. Correction of the mutation should result in proper CFTR glycosylation, detectable by western blot. To generate lung cells, we utilized both spontaneous and more directed *in vitro* differentiation.

Chapter 2: Materials and Methods

2.1 Cell lines and culture

Through Coriell Repository (Cambridge), we obtained CF primary fibroblast line GM04320 from a 17 year old male reported to be homozygous ΔF508 CFTR and showing pancreatic insufficiencies as well as advanced pulmonary disease. The CFTR mutation was interrogated by PCR amplification and sequencing (TABLE 1). CF fibroblasts and murine embryonic fibroblasts (MEFs; CF-1 mouse strain, Charles River) were maintained in DMEM supplemented with 10% FBS, 2 mM non-essential amino acids (NEAA), and 2 mM L-glutamine (all from Invitrogen). Fibroblast media is herein denoted as FM.

Human embryonic stem cell (ESC) line WA09 and transgene-free iPSC lines RC202 and RC204, which were provided by Darrel Kotton (*63*), were maintained in ESC media as outlined by National Stem Cell Bank protocol SOP-CC-001C. The cells were fed daily with DMEM-F12 (Invitrogen) supplemented with 20% Knockout Serum Replacer (Invitrogen), 2 mM each of L-glutamine and NEAA, 55 μM 2-mercaptoethanol (Sigma), and 4 ng/ml basic fibroblast growth factor (bFGF; Invitrogen). The CF-iPSCs that were generated in our laboratory were similarly maintained but with 40 ng/ml bFGF. All cells were cultured, except where denoted, in media under humidified conditions at 37°C and 5% CO_2.

Routinely, ESCs and iPSCs were passaged in clumps every 5-6 days. To passage, the cells were washed twice with dPBS (+ calcium / + magnesium), incubated in 1 mg/ml Collagenase Type IV (Invitrogen) for 3-5 minutes (enough for the edges of the colonies to begin to flip upwards but not enough for the whole colony to lift off the plate), and

TABLE 1 List of Primers. Where applicable, expected amplicon length and annealing temperatures used are also indicated.

Purpose	Forward	Reverse	length (bp)	Annealing T
sequencing of Δ	CCTGAGCGTGATTTGATAATGACC	GGACGTTTGTCTCACTAATGAG	533	48.2
generation of pMXs-GFP	CCATGAATTCATGGGTGAGCAAGGGCG	CGTAAGCGGCCGCTTTACTTGTACAGC	741	56.7
total *OCT4* expression	CAGTGCCCGAAACCCACAC	GGAGACCCACGCACGCCTCAAA	161	
endogenous *OCT4* expression	TTCGCAAGCCCTCATTTC	CCATCACCTCCACCACCT	114	
total *NANOG* expressin	TCTCCAACATCCTGAACCTCA	TTGCTATTCTTCGGCCAGTT	106	
endogenous *NANOG* expression	TTGGAAGCTGCTGGGGAAG	GATGGGAGGAGGGGAGAGGA	193	
total *c-MYC* expression	AGCGACTCTGAGGAGGAACA	CTCTGACCTTTTGCCAGGAG	87	
endogenous *c-MYC* expression	GCTGCTTAGACGCTGGATTT	TAACGTTGAGGGGCATCG	73	
total *KLF4* expression	GGGAGAAGACACTGCGTCA	GGAAGCACTGGGGGAAGT	88	
endogenous *KLF4* expression	GACCACCTCGCCTTACACAT	TTCTGGCAGTGTGGGTCATA	71	
total *SOX2* expression	CAAGATGCACAACTCGGAGA	GCTTAGCCTCGTCGATGAAC	95	
endogenous *SOX2* expression	TGCTGCCTCTTTAAGACTAGG	CCTGGGGCTCAAACTTCTCT	75	
REX1	TCCTCTTGGCACCTGTGG	AGGCTGGCAGAGAAGTCG	112	
CRIPTO	AGATGGCCCGCTTCTGTA	GAGATGGACGAGCAATTCC	108	
GAPDH	GCTCTCTGCTCCTCCTGTTC	ACGACCAAATCCGTTGACTC	115	
Cel-1 assay	GATAATTGGAGGCAAGTGAAT	CTGGCACCATTAAAGAAAATAT	195	58
amplification of CF for donor	TCTTCCTCTGCTACCTCC	ATTTGTTCAAAGCCAGGG	1563	40-52
amplification of lox-P flanked selection marker	GGGTAACCTAGGATAACTTTCGTATAAATGTATGCTATA CGAAGTTATGGTAAACTCGATGTAAGCTTGCCGTTGTC	CATGAGTGGGAGGAATGAGCTGGCCCATAACTTCGT ATAATGTATGCTATACGAAGTTATCCTAGGGTGAATGG	2619	
5' analysis of TI (1/1')	AAATCAGTGCTTTTTCGAGGTTAGGAG	CGACTGTGCCTTCTAGTTGC	1785	55
3' analysis of TI (2/2')	CCTGCAGCCCGGGGGATCTAT	GCCATTCATGTAAGCATGCAACCC	923	58
distinguishing modified vs unmodified allele (3/3')	CAGTGCTTTTTCGAGGTTAGG	GATGCAGAGGCAAAATGAAGA	4487 or 1928	
distinguishing modified vs unmodified allele (4/4')	AAATCAGTGCTTTTTCGAGGTTAGGAG	TTCTCTGCTGGCAGATCAATGCTCA	4369 or 1835	
RT-PCR Exon 8-9	ACGACGGAATTTGGGGAATTA	CTTGCTCGTTGACCTCCACT	460	
RT-PCR Exon 9-13	AATGGTGATGACAGCCTCTTC	GGAGACAGGAGCATCTCCTTCTAA	768	
RT-PCR *SOX17*	AAGGGCGAGTCCCGTATC	TTGTAGTTGGGGTGGTCCTG	221	
RT-PCR *GAPDH*	TCTTTTGCGTCGCCAGCCGA	CCUGCAAAUGAGCCCCAGCC	379	
qRT-PCR TaqMan *CFTR*	catalog #: Hs00357011_m1			
qRT-PCR TaqMan *NKX2.1*	catalog #: Hs00968940_m1			
qRT-PCR TaqMan *SOX9*	catalog #: Hs00165814_m1			
qRT-PCR TaqMan *P63*	catalog #: Hs00978343_m1			
qRT-PCR TaqMan *FOXP2*	catalog #: Hs00362818_m1			
qRT-PCR TaqMan *OCT4*	catalog #: Hs00999634_gH			
qRT-PCR TaqMan *FOXA2*	catalog #: Hs00232764_m1			
qRT-PCR TaqMan *PAX8*	catalog #: Hs00247586_m1			

then gently washed twice with dPBS. Using a 5 ml pipette tip, the cells were scraped to cut into pieces and then gently pipetted up and down. The cells were split 1:12 – 1:6. The cells were plated on 0.1% gelatin-coated (Sigma) plates pre-seeded (24 hours) with irradiated MEFs (iMEFs) as a feeder layer at ~ $2x10^5$ cells/well in a 6-well plate. For certain applications, ESCs or iPSCs were maintained on Matrigel. In such cases, plates were coated with 30% Matrigel (Becton Dickinson) at 4°C overnight or for 3+ hours at 37°C. The ESCs or iPSCs were treated with dispase (STEMCELL Technologies), manually broken into pieces, and split 1:3 or 1:4.

2.2 iPS Generation

pMXs-based retroviruses expressing reprogramming factors OCT4, SOX2, KLF4, c-MYC, and NANOG was provided by K. Plath and obtained through Addgene (*56*). To generate pMXs-GFP, the ORF for GFP was amplified by PfuUltra Hot Start DNA Polymerase AD (Agilent Technologies) from CCR5-pgk-gfp with primers (TABLE 1) containing novel EcoRI and NotI restriction enzyme sites. This amplicon was digested and cloned into the pMXs backbone, obtained by EcoRI and NotI digestion of pMXs-KLF4. The protocol for VSV-G pseudotyped retroviral vectors was adopted from Park and Daley (*120*). For each reprogramming factor, two 150 mm plates were seeded with approximately 7 million Plat-GFP Retroviral Packaging Cell Line (Cell Biolabs, Inc.). The next day, each plate of Plat-GFP cells was transfected with 6.25 μg of the pMX retroviral vector and 0.9 μg pCMV-VSV-G (provided by B. Weinberg and obtained through Addgene) (*121*) in a total volume of 30 ml media (DMEM, 10% FBS, 10 μg/ml blasticidin) containing 50 μl Fugene 6. Virus-containing media (a total of 60 ml per vector) was collected 48 hours after transfection, filtered through a 0.45 μm filter into a

polyallomer centrifuge tube, and spun down at 26K rpm for 2 hours (in a SW288 Beckman rotor). The pelleted virus was then resolubilized in 600 μl DMEM/F12 (for a 100x concentration). The virus titer was determined by infecting primary human fibroblasts with pMXs-GFP (simultaneously packaged and concentrated) and assessing the frequency of GFP-expressing cells by flow cytometry.

To reprogram fibroblasts (summarized in FIGURE 2.1), cells were plated at 10^5 cells per well of a 6-well plate on day 0. On days 1 and 2, the cells were transduced with all five reprogramming factors by spinfection (200 g for 30 minutes) at a multiplicity of infection of 21.5, in FM supplemented with protamine sulfate (10 μg/ml). Fifty-six hours after the initial transduction, the transduced cells were trypsinized and transferred (1:1) onto gelatin-coated plates previously seeded with iMEFs; 24 hours later, the media was switched from fibroblast media to ESC media containing 40 ng/ml bFGF, as indicated above. Starting on day 12, the cells were fed with hESC media that had been pre-conditioned on iMEFs. The cells were re-fed daily.

Approximately 16 days after transduction, we began to identify individual iPSC-like colonies based on morphology. Live-cell staining with anti-TRA-1-60 monoclonal antibody (mouse IgM; 1:200; Stemgent) followed by Alexa 488 goat anti-mouse IgM (1:250; Invitrogen) was used to identify reprogrammed colonies. When TRA-1-60 expressing colonies (Tra-1-60$^+$) were of sufficient size (generally, taking up at least ¾ of the field when looked under the microscope at 10x magnification), each clone was manually cut into 4-9 pieces, depending on size, with a glass, end-closed pipette. The pieces were then collected and seeded into a well of a 12-well plate (one colony per well) pre-seeded with iMEFs and ESC media supplemented with 10 uM Y-27632 (Alexis

Biochemicals). Depending on recovery rate and growth, these could then be subsequently split 1:2 to 1:4 at next passage. In total, we clonally picked 32 colonies that were TRA-1-60^{+}, nine of which were expanded and cryopreserved. Based on colony growth and morphology, we selected two clones (iPS 17 and 28) for more extensive characterization.

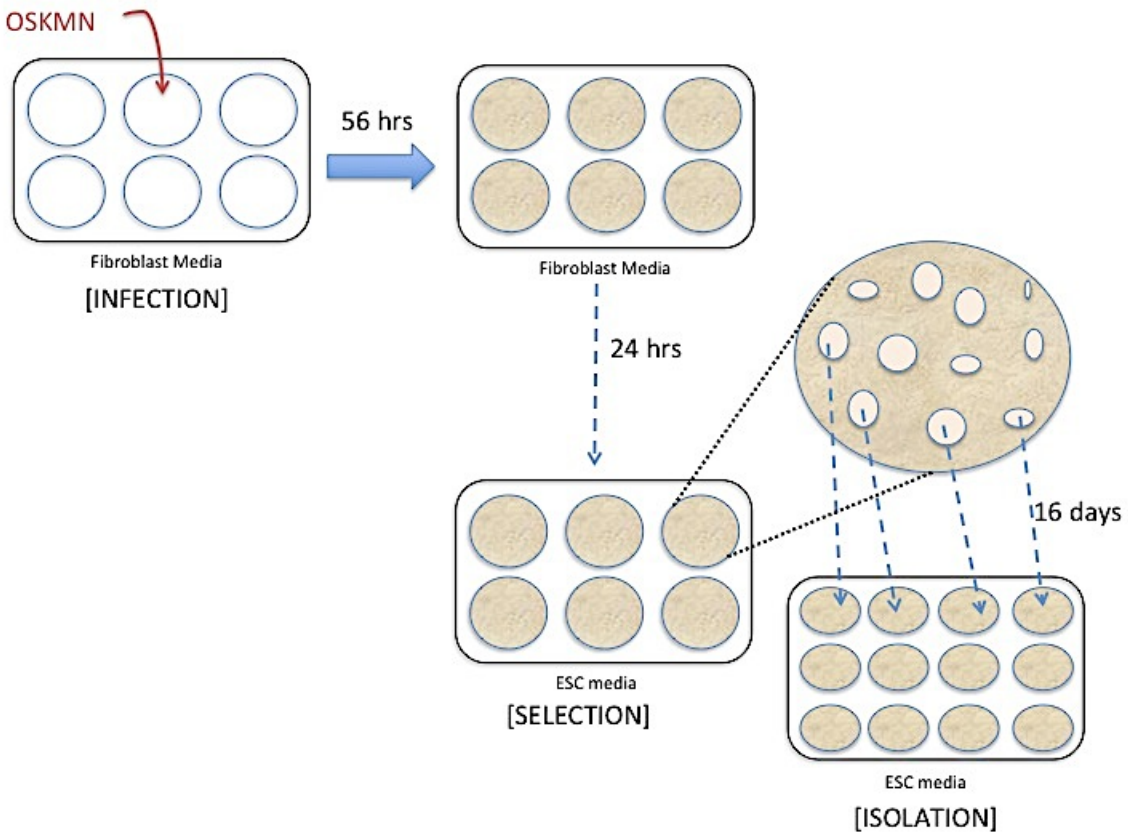
2.3 iPS Characterization

In addition to live-cell staining for TRA-1-60, potential iPSC clones were also characterized via live-cell staining for TRA-1-81 (mouse IgM; 1:200; Millipore). Co-expression of OCT4 and SSEA4 was assessed according to the WiCell protocol SOP-CH-102B and analyzed by flow cytometry on an LSR-II running BD FACSDiva Software (Becton Dickinson). Generally, after excluding CD29^{+} MEFs, the presence of $\geq$ 90% OCT4^{+}/SSEA4^{+} double positive cells (of total population excluding MEFs) was considered a relatively undifferentiated population with minimal spontaneous differentiation.

Alkaline phosphatase activity was accessed using VECTASTAIN ABC-AP Kit (Vector Laboratories). Briefly, after washing cells twice with DPBS, the cells were fixed for at least 30 minutes with 2% paraformaldehyde in DPBS. The paraformaldehyde was removed and replaced with 2 ml per well of a 6-well plate with substrate solution 123 for 30 minutes to 2 hours in the dark, per manufacturer's protocol. The cells were then washed with DPBS containing 0.1 Tween-20 for 5 minutes and rinsed with DPBS.

Karyotyping of CF-iPSC clones was performed at Texas Children's Hospital Clinical and Research Cytogenetic Laboratory, where they evaluated approximately 20 metaphase cells per cell line.

FIGURE 2.1 Reprogramming of CF-specific fibroblasts. Human fibroblast cells derived from a patient with CF were transduced with retrovirus to deliver the reprogramming cocktail (OCT4, SOX2, KLF4, c-MYC, and NANOG) in fibroblast media. After 56 hours, the cells were transferred onto gelatin-coated plates pre-seeded with irradiated mouse embryonic fibroblasts (iMEFs; shown as light brown), still in fibroblast media, and at sufficiently low density to enable clonal growth of individual cells surviving the infection and transfer. Twenty-four hours after transfer, when the transduced cells plated, cells undergoing reprogramming were selected for in ESC media; 16 days later, colonies arising out of the selection were clonally picked and transferred to iMEF-seeded 12-well plates for further expansion and characterization.



To assess the ability to generate teratomas, iPSCs were first cultured on 30% Matrigel and fed daily with mTESR (STEMCELL Technologies) to remove iMEFs. Between 2 and 3 million iPSCs were injected into the kidney capsule or testis of six-weeks old Fox Chase SCID beige mice (Charles River) and observed weekly for tumor growth. This was performed on 3 mice per cell line. The tumors were removed around 6 – 8 weeks after injection, paraffin embedded, prepared for histological examination by hematoxylin and eosin, and analyzed (Applied Stem Cell Inc.).

Qualitative RT-PCR was used to assess if the retroviral transgenes used in reprogramming had been silenced. To do this, total RNA was isolated from iPSCs or WA09 cells (for comparison) with the RNeasy Kit (Qiagen) and cDNA synthesized using Improm-II Reverse Transcription System (Promega). Levels of expression of reprogramming genes were determined by SYBR Green (Applied Biosystems) incorporation using primers (TABLE 1) designed to distinguish endogenous (one primer of each set includes 3' UTR) from total (both primers in coding regions) expression of each reprogramming gene. The RT-PCR was performed in two replicate samples for each cell line in a 96-well format on the ABI 7900, and expression levels were normalized to that of GAPDH.

Additionally, mRNA expression levels of 44 human ESC-specific genes in our CF-iPSCs were assessed by qRT-PCR using RT^2 Profiler PCR Array Human Embryonic Stem Cells (SABiosciences) per manufacturer's protocol. The array also included Beta-2-microglobulin, Hypoxanthine phosphoribosytransferase 1, Ribosomal protein L13a, Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and beta-Actin. For comparison,

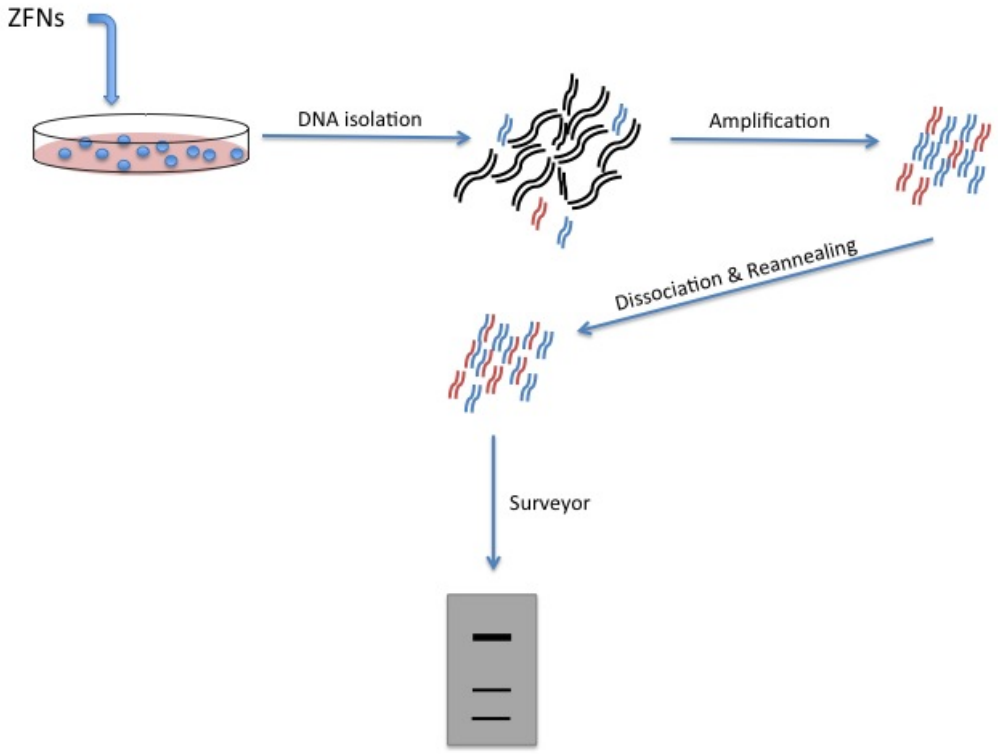
expression levels in human ESC line WA09 was simultaneously assessed. Analysis was performed using the data analysis template provided by the manufacturer.

2.4 Correction of CF-iPSCs

A potential ZFN target in exon 10 upstream of the ΔF508 mutations was identified and ZFNs were bioengineered by Sangamo Biosciences to bind and create a DSB. The efficacy of cutting at the targeted loci by ZFN pair 12987 (recognizing TAGAAGtGAAGTCTGG) and 9940 (ATTATGGGAGAACTG) in the form of expression plasmid or *in situ* generated RNA was determined by Cel-1/Surveyor assay (*91*, *93*). To generate the RNA for each ZFN, 1 μg of *XbaI* linearized plasmid DNA was used as per protocol to generate capped RNA using MessageMAX™T7 ARCA-Capped Message Transcription Kit (CellScript Inc.). This was then treated with DNaseI, a poly(A) tail was added by Poly(A) Polymerase Tailing Kit (Epicentre), and the resulting product was purified by MegaClear Kit (Ambion). The size, quality, and concentration of *in vitro* transcribed RNA was determined by Agilent Bioanalyzer.

The Cel-1 assay takes advantage of the infidelity of non-homologous end joining as the default repair mechanism for DSBs in the absence of a homologous template or donor (FIGURE 2.2). Two million dissociated iPSCs are first transfected with equal amounts of each ZFN and plated onto gelatin-coated plates seeded with iMEFs and incubated at either 37°C. Two days later, total DNA was isolated from the transfected cells and the 195 bp around the ZFN-targeted region was PCR amplified (TABLE 1). After denaturation and reannealing of the amplicon (95°C for 10 minutes; cool to 85°C at -2°C/sec; cool to 25°C at -1°C/sec; hold at 4°C), which results in a mixture of homoduplexes (of either two strands with the same NHEJ-induced mutation or two

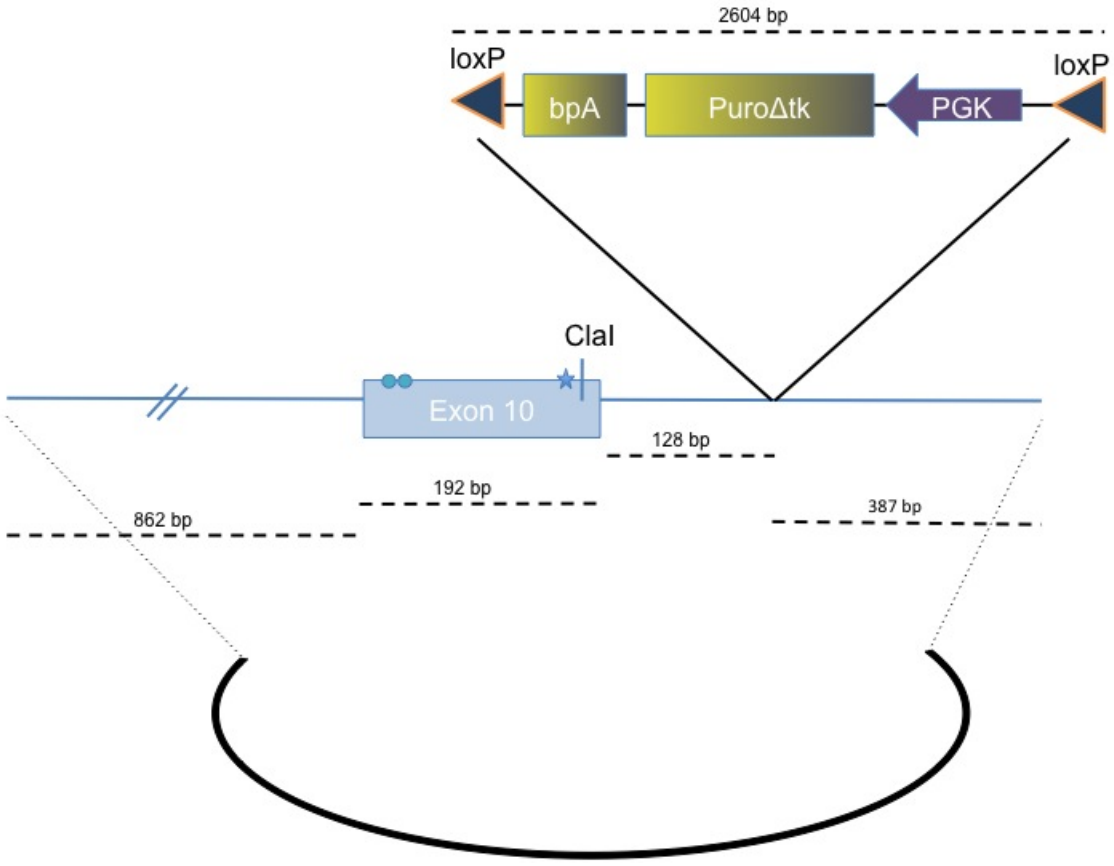
FIGURE 2.2 Efficacy of ZFN cutting can be determined by Cel-I analysis. To assess efficiency of cutting by ZFNs at the target loci, equal amounts of each ZFN are delivered to dissociated induced pluripotent stem cells. DNA is isolated from the cells, some of which was cut by the ZFNs and subsequently repaired by mutation-inducing NHEJ (red) while others were not cut by the ZFNs and thus wt (blue). The specific target is amplified and the amplicon is dissociated and re-annealed. This results in a mixture of duplexes with two wt strands, two mutated strands, or one mutated and one wt strand. The latter heteroduplex is recognized and leaved by Surveyor nuclease. By resolving undigested and digested material, efficacy can be quantified as a measurement of the ratio of cut band to total band intensity.



strands that were not cut by the ZFNs and thus wt) and heteroduplexes of NHEJ-induced mutated and ZFN-uncut strands. Surveyor nuclease was added to the mix to cut any heteroduplexes. This was then resolved on 10% TBE gels at 100V for 1 hour. Surveyor nuclease-digested heteroduplexes would result in two bands, 110 and 85 bp, quantifiable using a phosphoimager. The efficacy of cutting is a measurement of the relative ratio of digested heteroduplexes to total amount of undigested and digested as determined by the formula: (1-√(uncut/total)) x 100.

The donor plasmid (FIGURE 2.3) used to facilitate HDR was generated by amplifying from BAC gDNA (RP11-1152A23) approximately 1.6 kb *CFTR* sequences flanking the intended ZFN cut site (TABLE 1). This amplicon was cloned into the pSC-B as per StrataClone Blunt PCR Cloning Kit (Stratagene). An AvrII site was introduced via Quikchange II Site-Directed Mutagenesis (Agilent Technologies) in intron 10, 128 bases -downstream of the end of exon 10. By Quikchange Multi Site-Directed Mutagenesis (Agilent Technologies), two silent mutations were introduced in the sequences recognized by the right ZFN and another silent mutation was introduced 22 bases downstream of the sequences restoring F508. This latter silent mutation created a novel ClaI site. A selection marker conferring phosphoglycerate kinase-driven puromycin resistance and thymidine kinase sensitivity (pgk-puro-TK) was amplified from pPthc-Oct3/4 (received from Dr. Naoki Nakayama) (*122*) using primers (TABLE 1) containing loxP sites and flanked by AvrII sites and Phusion Hot Start II high fidelity DNA polymerase (New England Biolabs) in a way such that the orientation of the selection cases was antisense to *CFTR*. This amplicon was cloned into the pSC-B backbone

FIGURE 2.3 Donor for homology directed repair of CF. Approximately 1.6 kb of wild-type sequences for *CFTR* was used as homology arms and a 2.6 kb selectable marker flanked by loxP sites was cloned into intron 10, downstream of exon 10 splice donor sites. The selectable marker was cloned in such that expression is antisense to that of *CFTR*. To disrupt recognition (and cutting) of the donor by the co-delivered ZFNs, two silent mutations were introduced in the sequences that would be normally recognized by ZFN-R (circles). A third change was made to the wild-type sequences: a ClaI site was introduced via silent mutation approximately 20 bases downstream of the wild-type sequences to correct ΔI507/ΔF508 (star).



containing *CFTR* homology sequences at the AvrII site by In-Fusion Cloning (Clonetech Laboratories). The donor plasmid was confirmed by sequencing.

Delivery of ZFNs and/or donor plasmid to cells was accomplished via nucleofection using Human Stem Cell Nucleofector Kit 1 (Lonza), adapting manufacturer protocol. For each condition, two million human ESCs were dissociated into single cell suspension via Accutase (STEMCELL Technologies) treatment. Incubating the iPSCs in Accutase for 2-3 minutes at 37°C results in most of the MEFs lifting off the plate while the iPSC colonies remained loosely attached with cells rounding up, making them easily dissociated by gentle pipetting with PBS after the Accutase was removed. The cells were centrifuged in microfuge tubes at 200g for 5 minutes and the PBS aspirated. The cells were then resuspended in 100 µl of the supplemented nucleofection solution. This was then added to another microfuge tube containing the various DNAs/RNAs combinations: 1 or 2 ug of each ZFN DNA expression plasmid or 1.5 or 3 ug of each ZFN RNA plus 4 or 8 ug of the donor (TABLE 2). Controls for ZFN cutting alone did not include donor; controls for random integration of donor sequences, in the absence of ZFN-mediated cutting, included only donor without ZFNs. This mixture was then placed into the cuvettes provided and nucleofected using the Amaxa program A23. Subsequently, 500 ml of ESC media containing the appropriate amount of bFGF and supplemented with 10 uM Y-27632 was gently added to the nucleofected mixture, which was then split between two wells of a 6-well plate pre-seeded with puromycin-resistant iMEFs in supplemented ESC media. After 24 hours, the cells were fed daily. Puromycin selection (0.5 ug/ml) began 4 days post nucleofection and, 2-5 days later, puro-resistant colonies were clonally picked

TABLE 2 Matrix of correction experiments.

	Donor (µg)	ZFN Plasmid (µg)	ZFN RNA (µg)	# puro^R clones picked	# puro^R clones analyzed*
Experiment 1					
	4	2		23	11
	4		1.5	13	5
Total				**36**	**16**
Experiment 2					
	4			7	5
	4	2		5	4
	4	1		7	6
	8	2		6	2
	8	1		14	7
	4		3	3	3
	4		1.5	6	5
	8		3	4	3
	8		1.5	16	14
Total				**68**	**49**

* After picking, some of the clones were contaminated and removed from further analysis.

in the same manner as isolation of potential reprogrammed cells and expanded (still under puromycin selection, generally, for 3-4 passages).

2.5 Analysis of targeted clones

Targeted integration of the donor was assessed first by PCR using two sets of primers (TABLE 1). One primer of each set recognizes *CFTR* sequences outside the homology arms found in the donor and the other primer of each set recognizes sequences in the drug selection cassette. Thus, the primers were designed such that amplification would occur only in the case of integration of the donor at this locus. The 5' amplicon was produced using Expand Long Template PCR (Roche), and the 3' amplicon was produced using GoTaq Green Master Mix (Promega). To confirm correction of the mutation, the 5' amplicon was sequenced. To determine which allele was uncorrected, the unmodified allele was preferentially amplified using Expand Long Template PCR and primers that recognized *CFTR* sequences outside the donor homology arms so that the unmodified allele amplicon was ~1.9 kb while the modified allele amplicon was ~4.5 kb (TABLE 1). The unmodified allele amplicon was gel isolated via Nucleospin Gel and PCR Cleanup (Macherey Nagel) and sequenced.

To assess expression of the corrected *CFTR* mRNA, RNA was isolated with RNeasy kit (Qiagen) and cDNA synthesized using Improm-II oligo dT kit (Promega). The *CFTR* cDNA was amplified from either the border of exon 8-9 to exon 11 or from exon 9 to exon 13 (TABLE 1) and subsequently sequenced.

Southern blotting was used to assess site-specific integration of the donor. A radio-labeled DNA probe was generated by digesting the donor plasmid with NdeI and SpeI and separated on 0.8% agarose gel. A 2.3 kb fragment containing pgk-puro-TK

sequences was gel purified and labeled with [alpha-^{32}P]dCTP using Prime-It II Random Primer Labeling kit (Agilent Technologies) per manufacturer's instruction. Genomic DNA (25 mg), digested with SpeI overnight and purified by phenol/chloroform extraction, was resolved on 1% agarose gel, transferred to a Nytran Super Charge membrane (Schleicher and Schuell), and hybridized with ^{32}P-labeled probe. The membrane was exposed and image scanned using a phosphorimager system (Molecular Dynamics).

2.6 Cre-mediated excision of selectable marker

Excision of the loxP-flanked selectable marker was accomplished via transient expression of Cre-recombinase. Four μg of either pBS513 EF1alpha-cre [Addgene 11918 provided by B. Sauer] (*123*) or pCAG-Cre [Addgene 13775 provided by C. Cepko] (*124*) was delivered to 2 million corrected iPSCs (previously dissociated into a single cell suspension via Accutase) by nucleofection as described previously in Section 2.4 and plated onto irradiated MEFs. Successful excision was determined by either FIAU resistance (1 μM; Moravek Biochemicals) or by first picking and expanding individual clones in replicates and identifying puromycin sensitive clones. Clones that were FIAU resistant and/or puromycin sensitive were tested molecularly as confirmation of excision of selectable markers using the same primers used to determine targeted integration in Section 2.5. Cre-excision was confirmed molecularly by PCR using primers that anneal to sequences outside donor homology, yielding three possible amplicons: 1) a 4.4 kb amplicon corresponding to the modified allele, 2) a 1.8 kb amplicon to the modified allele after excision, and 3) a 1.8 kb amplicon to the unmodified allele. The modified allele post excision was distinguished from the unmodified allele by ClaI digestion.

2.7 *In vitro* differentiation

For differentiation into definitive endoderm, precursor to the epithelial lining of the forgut, we adapted the protocol of D'Amour et al. (*125*). To induce differentiation, iPSCs grown on iMEFs were cultured in RPMI supplemented with 100 ng/ml Activin A (R&D Systems) and varying concentrations of fetal bovine serum (0% on day 0, 0.2% on day 1, and 2.0% every day thereafter). RNA was isolated at various time points and analyzed by RT-PCR for expression of *CFTR* (primers spanning exons 9 through 13) and *SOX17* (TABLE 1), a marker of definitive endoderm.

In 2009, Van Haute et al. described the differentiation of human ESCs into lung epithelial-like tissue that showed increased *FOXJ1* and β-*tubulin* mRNA expression and evidence for ATII expression via immunostaining (*126*). Trying to replicate their findings, we took two iPSC lines that had been corrected and the uncorrected iPSC line and followed their protocol. Cells were collagenase treated for 7 minutes so that the colonies rounded up but still remained loosely attached to the surface of the plate. We then seeded 30 clumps, defined either as 30 rounded colonies or 30 pieces cut from colonies, onto 12 mm hydrophilic PTFE Millicell cell culture inserts (0.4 μm; EMD Millipore) that were placed into gelatin and iMEF-coated 12-well plates with 600 ml ESC media outside the insert and 400 ml inside the insert. After 24 hours, the ESC media inside and outside the insert (liquid-liquid conditions) was changed daily. Four days after seeding the clumps, the media was switched to differentiation media (ESC media without β-mercaptoethanol or bFGF). Media was changed daily for four more days still in liquid-liquid conditions, after which the cells were then cultured in air-liquid conditions with differentiation media only outside the insert. After 15 days, RNA was isolated and expression of *GAPDH*, *SOX17*, *CFTR*, and *FOXJ1* was assessed by RT-PCR.

To establish *NKX2.1*-expressing progenitor cells, which give rise to lung and thyroid, we followed the directed differentiation protocol as described by Green et al. and Longmire et al. (*79*, *127*). Prior to differentiation, the iPSCs were first adapted to single-cell dissociation via TrypLE (Invitrogen) and cultured on Matrigel in mTeSR. To induce differentiation, embryoid bodies are first formed: the cells are treated with dispase until the edges of the colonies have rounded up, washed three times with IMDM (Invitrogen), and the colonies were broken into clumps of 10-20 cells in pre-warmed serum-free differentiation media (SFD) (*128*) supplemented with 10 μM Y-27632, 3 ng/μl BMP4, and 10 ng/μl Wnt3a. These were then transferred to low attachment plates and incubated at 37°C, 5% CO_2, and 5% O_2 for 24 hours. Between days 1 and 4, the cells are fed every other day with induction media (SFD supplemented with 10 μM Y-27632, 0.5 ng/μl BMP4, 2.5 ng/μl FGF2 and 100 ng/μl Activin A). On day 4, expression of endodermal surface markers c-KIT and CXCR4 was assessed by flow cytometry (antibodies from Invitrogen). If we obtained populations of >70% double positive, we then proceeded with anteriorization; 60,000 to 80,000 single cells were plated into each well of a gelatin-coated 24-well plate and cultured in SFD supplemented with 10 μM Y-27632, 200 ng/ml Noggin, and 100 ng/ml SB-431542. From day 6 until day 19, the cells were cultured in prewarmed ventralization media (SFD supplemented with 500 ng/ml FGF2, 100 ng/ml Wnt3a, 10 ng/ml FGF7, 10 ng/ml BMP4, 20 ng/ml EGF, and 100 nM Retinoic Acid). All reagents, unless noted otherwise, were obtained from R&D Systems. At indicated time points, cells were pelleted and stored for future protein or RNA isolation and analysis. The expression profile of differentiated cells was measured in triplicate using TaqMan probes from Life Technologies. Gene expression was normalized against housekeeping

gene 18S and compared to undifferentiated samples (TABLE 1). Total RNA from fetal lung and fetal thyroid samples (BioChain) was used for comparison.

To detect CFTR protein, samples from day 19 of differentiation, along with Calu-3 cells as a positive control, were lysed with 1% Triton X-100, 50 mM Tris HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, and a protease inhibitor mixture (Roche) while on ice and centrifuged for 15 minutes at 6,000 g. After cell debris was removed, protein amount was quantified by BCA protein assay (Thermo Fisher). Total protein from differentiated samples (100 μg) and Calu-3 cells (10 μg) were treated with denaturation buffer (as described by New England Biolabs' protocol for PNGase F) with or without PNGase F for 1 hour at 37°C. The reaction was stopped by adding 4x Laemmli sample buffer (Bio-Rad) to the mixture. The protein (50 μg from differentiated samples and 5 μg from Calu-3 samples) was then separated by NuPAGE SDS PAGE Gel System (Life Technologies) on 7% Tris-Acetate gel (Life Technologies), transferred onto Hybond-C nitrocellulose transfer membrane (Amersham Biosciences). The membrane was blocked in 5% milk for 1 hour at room temperature and then incubated in mIgG2b 596 (1:1000; Cystic Fibrosis Foundation Therapeutics) overnight at 4°C. After washing with TBE and 0.1% Tween-20, the membrane was incubated in secondary HRP-linked anti-mouse IgG (1:5000; Cell Signaling Technologies) for 1 hour at room temperature before detection by chemiluminescense using Amersham ECL Prime. Anti-Calnexin polyclonal antibody (1:300; Abcam) along with secondary anti-rabbit IgG IRDye 800CW (LI-COR) was used as a loading control. Fluorescence intensities of the protein bands were measured with an Odyssey Quantitative Infrared Imaging System.

Differentiated cells were also immunostained to visualize FOXA2 and NKX2-1. Cells were fixed with 4% paraformaldehyde for 30 minutes and permeabilized with 0.2% Triton X-100 for another 15 minutes. The sample was then blocked in 4% donkey serum for 30 minutes and incubated in primary antibody (FOXA2 goat polyclonal antibody, 1:100, Santa Cruz Biotechnology; NKX2.1 rabbit polyclonal antibody, 1:100, Abcam) at 4 degrees overnight. The cells were then washed twice and incubated in secondary antibody (1:100) at room temperature for 2 hours. Cy3 donkey anti-rabbit (Jackson Labs) and goat anti-mouse Alexa 488 (Life Technologies) were used as secondary antibodies.

2.8 Genome sequencing and analysis

Complete genome and whole exome sequencing was performed and analyzed as previously described (to be published)

> Human iPSCs were first depleted of MEFs by passage on matrigel in MEF-conditioned media. Genomic DNA was extracted with the QIAprep Spin Miniprep Kit. DNA sequences were obtained as described previously using DNA nanoball amplification and combinatorial probe-anchor ligation sequencing (Drmanac et al. 2010; Lee et al. 2010). Possible de novo non-synonymous coding variants (NSCV) were analyzed and filtered based on two independent criteria (somaticScore and refScore) and then combined to generate the list in Table S2. Since each cell line was compared to the reference genome (GRCh37/hg19) we used the calldiff function with somatic option in CGA Tools to generate pairwise comparisons of our four cell lines. All NSCV with SomaticScore greater than -10 from the 8 pairwise comparisons were combined with NSCV which had a refScore greater than 60 (indicating the likelihood of being homozygous reference) in the mutant CF fibroblast cell line. The compendium of putative NSCV was examined across all four cell lines by loading evidence file based BAM files for each cell line into the Integrative Genomics Viewer (Robinson et al. 2011). We considered false NSCV to consist of variations in areas with poor coverage (<9 reads), seen in the CF mutant fibroblast (Gore et al. 2011), variations present at <10% frequency, and those duplicating nearby repetitive sequences. The 10% frequency is based on observations of loci with variant reads present at 10% of total reads that were not called variant. Since the BAM files were generated from evidence files after de novo alignment, coverage was assessed indirectly based on calls in the masterVarbeta files; regions with

"no call" or "complex" in any of the cell lines. In other instances where there was no sequence data in the BAM files and a call of "hom ref", we assumed homozygous reference sequence when coverage scores were positive.

Exome Capture and Sequencing:

Two micrograms of genomic DNA was submitted to Axeq Technologies for human exome capture sequencing using TrueSeq 62 Mb target enrichment (http://www.illumina.com/documents/products/datasheets/datasheet_truseq_exome_enrichment_kit.pdf). Axeq Technologies performed sample validation, library preparation, exon enrichment, clustering and sequencing using illumina HiSeq 2000 Sequencer. Approximately 63,000,000 reads of an average size of 107 bp per sample was returned to us as two fastq files (one file per orientation).

Read Mapping:

Each pair of fastq files were aligned to human genome (hg19) using Novoalign (http://www.novocraft.com). All parameters were kept at the default settings, as recommended by Novocraft. SAMtools (http://samtools.sourceforge.net/) (Li et al. 2009) was used to sort the SAM files, create BAM files and generate their index files. Picard (http://picard.sourceforge.net/) was used to remove all the PCR duplicates from the BAM files. The Genome Analysis Toolkit (GATK) (McKenna et al. 2010) was used for local realignments, base quality recalibration, and variant calling. Parameters were set as described in GATK's Best Practices v3. GATK generated standard variant call format files (VCF http://www.1000genomes.org/wiki/Analysis/Variant%20Call%20Format/vcf-variant-call-format-version-41). The VCF files were annotated using snpEff (snpeff.sourceforge.net) (Cingolani et al. 2012) and ANNOVAR(Wang et al. 2010). From this point on we focused our analysis only on the putative coding non-synonymous variants.

Retroviral insertion sites were identified using methods derived from Complete Genomics based on assembling junctions. The reference genome, to which sequences were mapped, consisted of the GRCH37/gh19 bill augmented with two pieces of the long terminal repeats (LTR) common to all the retroviral vectors. These additional LTR sequences was added to the "end" of the genome, so that "left" and "right" sides of junctions could be defined.

The ZFN-Site Search Engine with the implemented FetchGWI source code (*129*) was used for *in silico* identification of closely related, homologous, off-target ZFN

recognition sites within two human reference genomes (GRCh37/hg19; NCBI36/hg18). Two bp mismatches per ZFN cassette and an allowed spacer region from 5 base pairs up to 16 base pairs were considered as stringency limits.

2.9 Acknowledgement of contributions

Dr. Ana Maria Crane performed reprogramming experiments, characterized and expanded a subset of the iPSCs, optimized Nucleofection of the iPSCs, performed all teratoma assays, did "Experiment 1" of correction experiments, and implemented the Activin-A differentiation of iPSCs. Analyses of Complete Genomic and whole exome sequencing were performed by the following: Drs. Manuel Gonzalez, Xuan Shirley Li and Shan Yang (the latter from Complete Genomics) performed most of the bioinformatics, and Drs. Philipp Kramer and Li created the analysis design and manually curated the findings. Differentiation of iPSCs into NKX2.1^{+} lung and thyroid progenitors was the collaborative work between Drs. Kramer and Dr. Finn Hawkins (the latter in Dr. Darrell Kotton's laboratory).

Chapter 3: Results

3.1 Generation of CF-iPSCs

Primary fibroblast cells, reported to be homozygous for the ΔF508 mutation, were obtained from Coriell Cell Repositories (GM04320 17yo male). Upon amplification and subsequent sequencing of the 500 bases surrounding the intended ZFN recognition sequence, we found GM04320 was actually compound heterozygous ΔI507/ΔF508 (FIGURE 3.1).

To reprogram the fibroblast cells, they were transduced with pMXs retrovirus to induce ectopic expression of reprogramming factors OCT4, SOX2, KLF4, c-MYC, and NANOG and transferred onto MEFs for subsequent culture. Reprogramming cells were selected for in ESC media. By day 16 after transduction, we observed colonies that resembled ES colonies morphologically; there was some variation in the extent of resemblance to ES colonies. Live-cell staining for TRA-1-60 was used to identify colonies for picking. Live-staining served as a useful tool in that it identified potential reprogrammed colonies even before they definitively exhibited ESC-like morphology (FIGURE 3.2).

Focus was placed upon two particular clones – designated iPS 17 and iPS 28 – based on growth properties and ESC-like morphology. Sequencing verified ΔI507/ΔF508 compound heterozygosity in the iPSCs (FIGURE 3.3a). Immunostaining for TRA-1-60, TRA-1-81, and Alkaline Phosphatase demonstrated that these cells expressed antigens characteristic of undifferentiated ESCs (FIGURE 3.3b). We also demonstrated that these cells can be passaged at length without significant spontaneous differentiation, as demonstrated in FIGURE 3.3.b via flow cytometric analysis of SSEA4 and OCT4

FIGURE 3.1 CF fibroblast sequencing. Primary CF fibroblast line GM04320 was obtained from Coriell Repositories. Sequencing revealed compound heterozygosity at the highlighted position (a mixture of bases T and A), indicating one allele is ΔI507 and the other ΔF508.

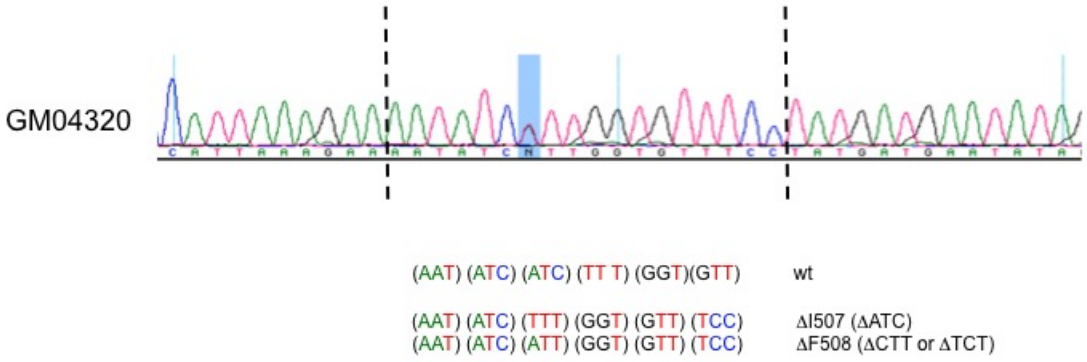


FIGURE 3.2 Isolating putative reprogrammed cells based on expression of TRA-1-60. Approximately 16 days after reprogramming, putative reprogrammed colonies were picked based on live-staining for stem cell markers TRA-1-60 rather than morphology. Each of these colonies were manually picked and transferred to one well of a 12-well plate coated with iMEFs. Staining and imaging performed 10 days post transductions.

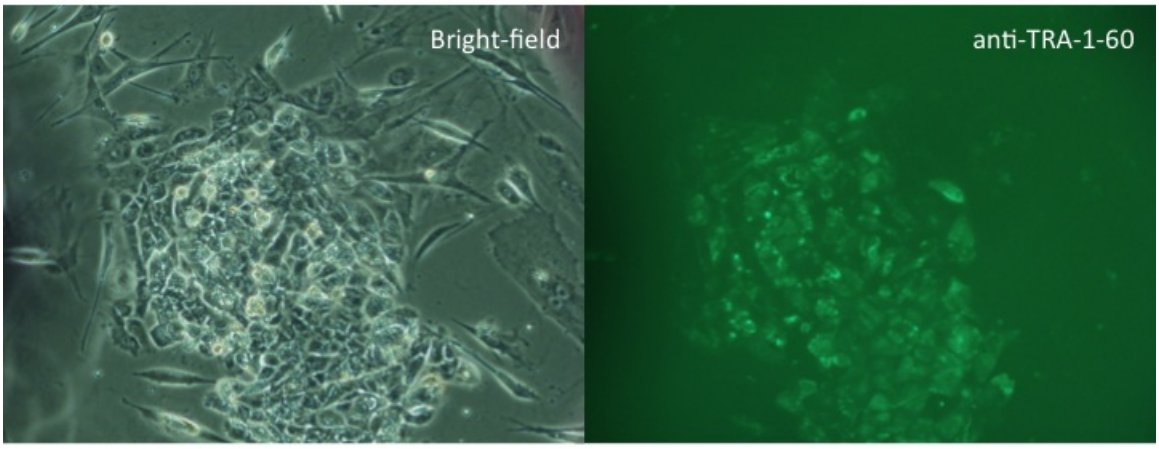
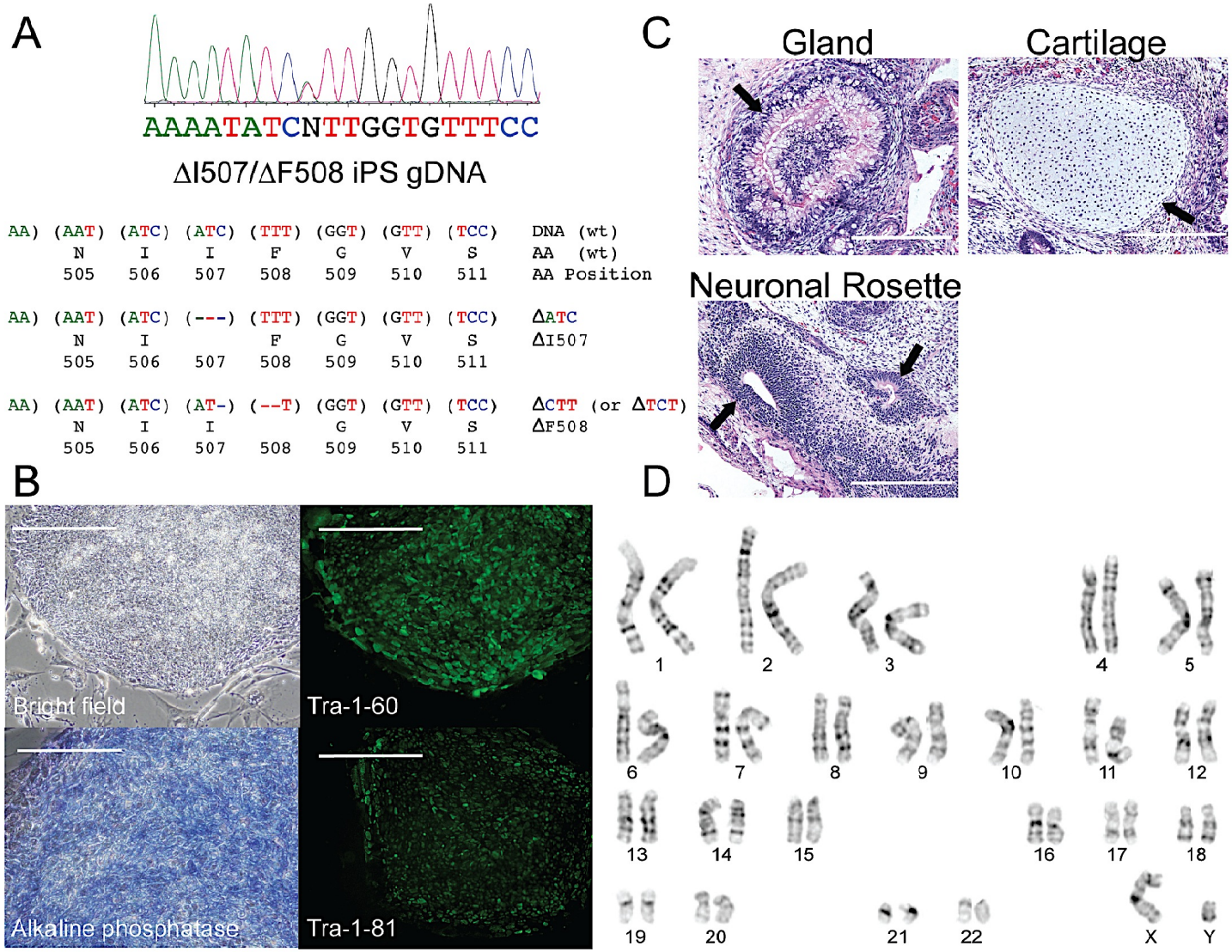


FIGURE 3.3 Characterization of iPSCs. a) Reprogrammed cells generated from fibroblast line GM04320 maintain ΔI507/ΔF508 heterozygosity. **b)** iPSCs express markers characteristic of embryonic stem cells. Shown here is iPSC clone 17 showing expression of surface markers TRA-1-60 and TRA-1-81, alkaline phosphatase, and co-expression of OCT4 and SSEA4 by flow cytometry. **c)** iPSCs injected into immune deficient mice give rise to teratomas with evidence of all three germ layers. **d)** iPSCs show normal male karyotype, suggesting reprogramming caused no gross chromosomal abnormalities.


A
AAAATATCNTTGGTGTTTCC
ΔI507/ΔF508 iPS gDNA
AA) (AAT) (ATC) (ATC) (TTT) (GGT) (GTT) (TCC) DNA (wt)
N I I F G V S AA (wt)
505 506 507 508 509 510 511 AA Position
AA) (AAT) (ATC) (---) (TTT) (GGT) (GTT) (TCC) ΔATC
N I F G V S ΔI507
505 506 507 508 509 510 511
AA) (AAT) (ATC) (AT-) (--T) (GGT) (GTT) (TCC) ΔCTT (or ΔTCT)
N I I G V S ΔF508
505 506 507 508 509 510 511
B
Bright field
Tra-1-60
Alkaline phosphatase
Tra-1-81
C
Gland
Cartilage
Neuronal Rosette
D
1 2 3 4 5
6 7 8 9 10 11 12
13 14 15 16 17 18
19 20 21 22 X Y

expression, both indicative of pluripotency, in iPS17 after 33 passages. Pluripotency of the iPSC was additionally established via teratoma assay. When iPSCs were injected into immune-deficient mice, recipient mice developed tumors consisting of cell types characteristic of all three germ layers: mesoderm (cartilage), ectoderm (neuronal rosettes), and endoderm (gland) (FIGURE 3.3c). Furthermore, the CF-iPSCs still had a normal male karyotype (FIGURE 3.3d).

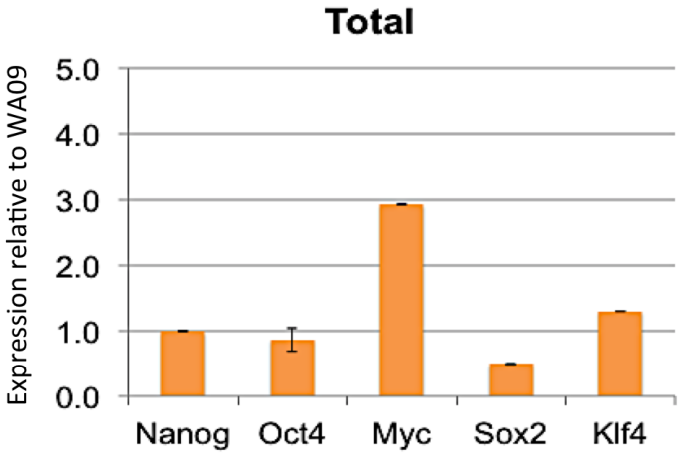
To evaluate silencing of reprogramming transgenes as well as activation of the human ESC pluripotency network, the total (FIGURE 3.4a) and endogenous (FIGURE 3.4b) expression levels of those reprogramming factors were measured by qRT-PCR and compared to expression levels in WA09. We did not observe significant difference between iPS 17 and WA09 in total and endogenous expression levels of any of the genes, consistent with transgene silencing. We also found little difference between iPS 17 and WA09 when comparing the expression profile of various genes implicated in maintenance of pluripotency and self-renewal of ESCs (FIGURE 3.4b).

3.2 Gene correction of CF-iPSCs

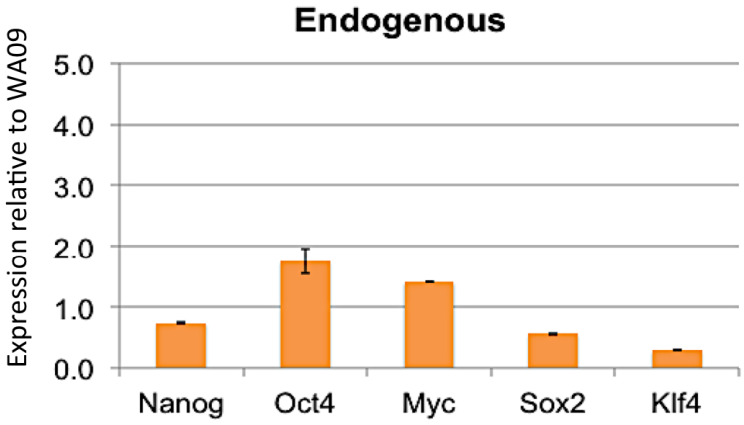
Correction of the CF-iPSC was accomplished via ZFN-mediated homology directed repair (a schema provided in FIGURE 3.5). The ZFNs were engineered to target exon 10 of the CF locus, creating a DSB approximately 110 bp upstream of either ΔI507 or ΔF508. The efficacy of ZFN cutting was first assessed in a Cel-1 assay by taking advantage of non-homologous end joining, an error-prone mechanism for default repair of DSBs when no appropriate donor template is available. Briefly, cells were nucleofected to deliver ZFN alone. Two days later, genomic DNA was isolated and the region immediately surrounding the target cut site was PCR amplified. The amplicon was

FIGURE 3.4 Transcriptional profile of CF-iPSCs. To determine whether reprograming transgenes were silenced, the levels of total (transgene + endogenous; **a**) and endogenous-only (**b)** RNA expression were measured by qRT-PCR for each of the reprogramming genes, normalized to expression of GAPDH and compared to levels of expression in WA09 human ESCs. No significant difference in total and endogenous levels between the CF-iPS and WA09 is consistent with transgene silencing; the greatest difference observed was an approximate 3-fold increase in total *C-MYC* expression levels in iPS 17 in comparison with WA09. **c)** A panel of 44 human ESC-specific genes was analyzed in WA09 and iPS 17 by quantitative RT-PCR. Again, iPS 17 showed a similar gene expression pattern compared to WA09. Outliers are indicated in red with the gene name. Consistently, we have observed ~100-fold increase in expression of *FGF5* in the WA09 cells relative to uncorrected iPS 17 or corrected iPS cells (Figure 3.16). Black lines indicate fold change of 1 and red lines indicate fold change of 3. Also included in the array were 5 housekeeping genes (*B2M*, *HPRT1*, *RPL13A*, *GAPDH*, and *ACTB*).

a


Total
Expression relative to WA09
5.0
4.0
3.0
2.0
1.0
0.0
Nanog
Oct4
Myc
Sox2
Klf4

b


Endogenous
Expression relative to WA09
5.0
4.0
3.0
2.0
1.0
0.0
Nanog
Oct4
Myc
Sox2
Klf4

c

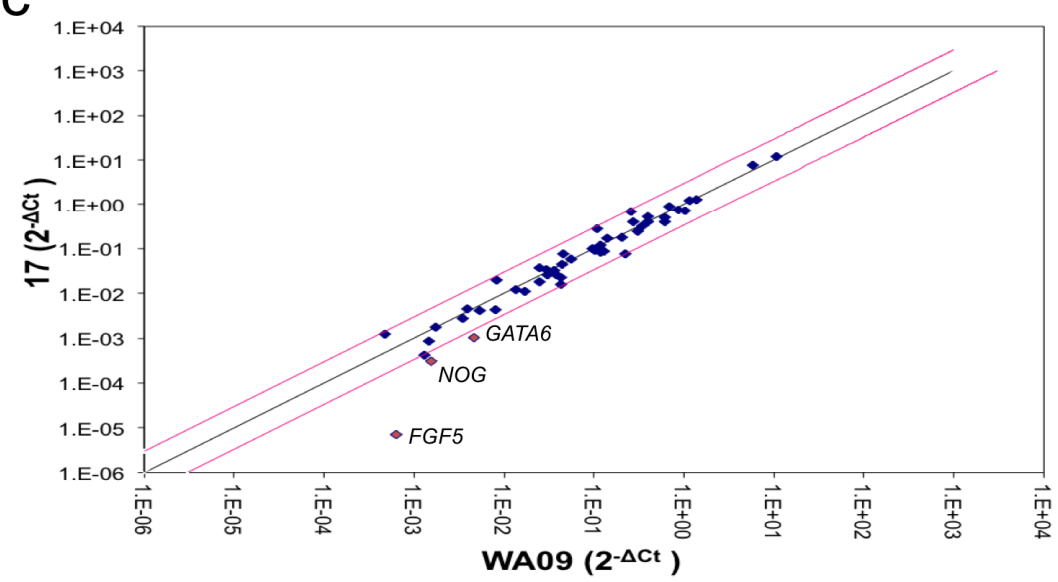
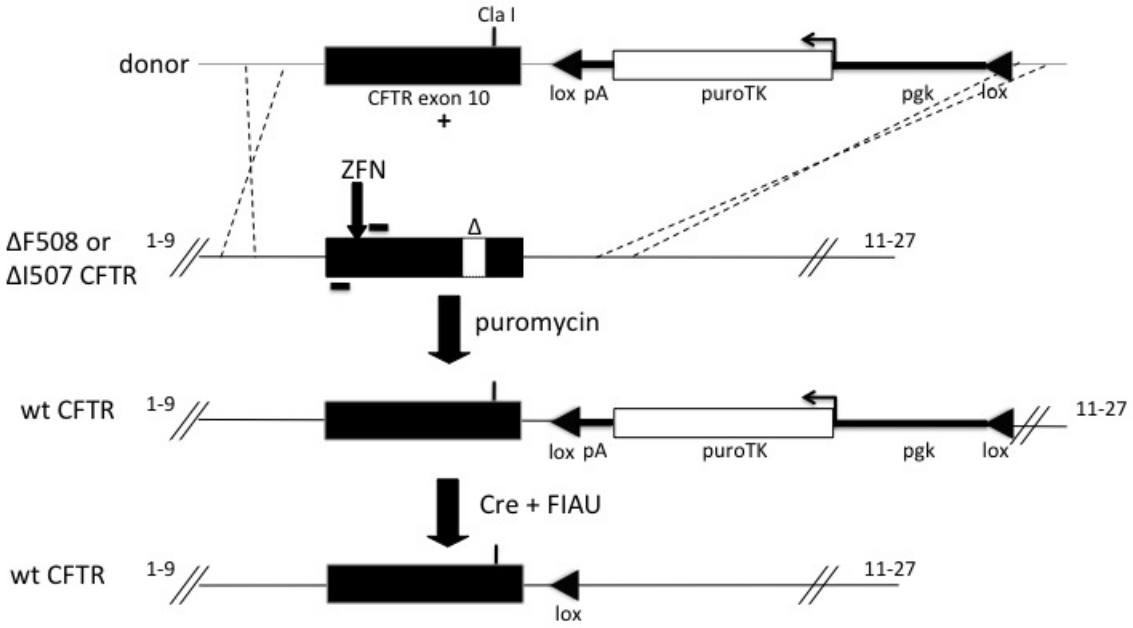

17 (2-ΔCt)
WA09 (2-ΔCt)
GATA6
NOG
FGF5

FIGURE 3.5 Schematic plan of proposed targeted CF correction. ZFNs, designed to induce a double strand break in exon 10 of the CF gene, are co-delivered with a donor molecule to serve as a template for homology directed repair of the DSB. Integration of the donor is selected for by puromycin selection. After successful targeted integration has been determined via molecular characterization, the selectable marker is excised by Cre-recombination of the loxP sites that flank the cassette, leaving one loxP site in intron 10.

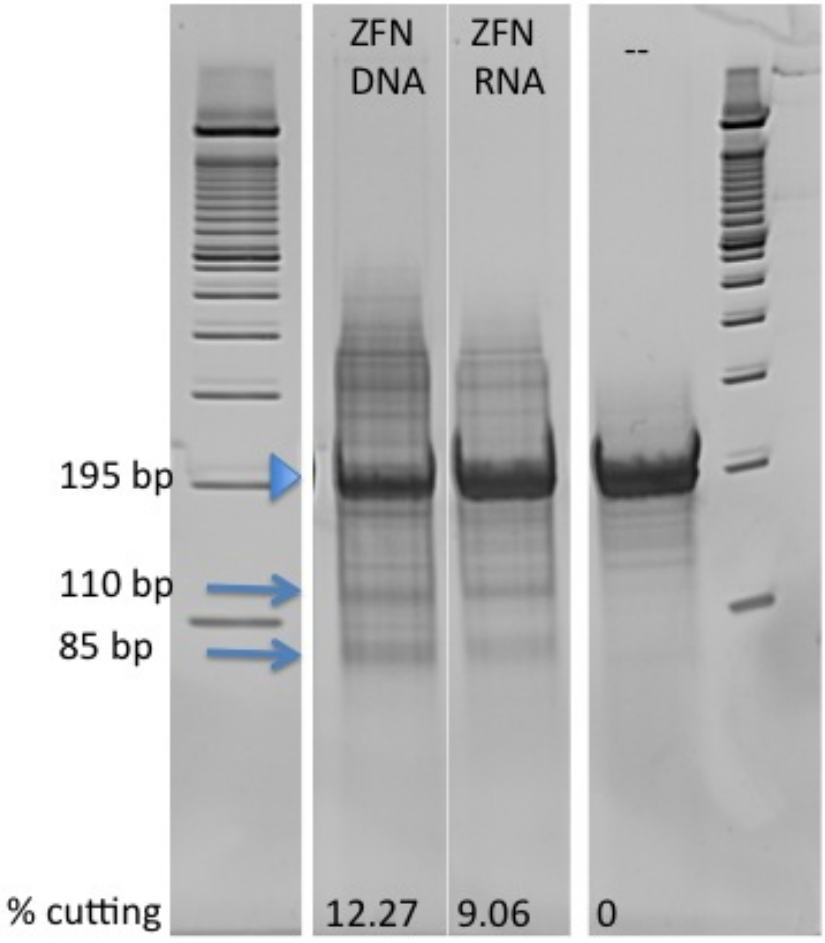


then denatured and renatured, treated with Surveyor nuclease, and the digested amplicon was resolved by gel electrophoresis. We found that 2 μg of each ZFN DNA expression plasmid or 1.5 ug of each ZFN RNA resulted in ~ 10% cutting, which in our limited experience was deemed sufficient to perform gene correction experiments (FIGURE 3.6).

For correction, accutase dissociated CF-iPSCs were nucleofected to co-deliver the ZFNs (either as DNA expression plasmids or *in vitro* transcribed RNA) along with a DNA donor plasmid to serve as a template for HDR (TABLE 2). This donor contained 1.6 kb sequences homologous to the genome flanking the ZFN cut site, which then includes all of exon 10 (see FIGURE 2.3). Several modifications were made to this homologous region. First, two silent bp substitutions were introduced in the sequences recognized by the ZFN-R to prevent cleavage of the donor by the ZFNs. Second, a novel ClaI site, which retained the exact coding sequence, was introduced approximately 20 bases downstream of the wild-type sequences to correct ΔI507/ΔF508. Lastly, a loxP flanked pgk-puro-TK selectable marker was inserted in the anti-sense direction approximately 125 bp downstream of exon 10, in intron 10 of the donor. This enabled us to select for those iPSCs that had integrated the donor into the genome.

Puromycin resistant colonies (each arising from a single cell since the iPSCs were dissociated and subsequently plated in a diluted manner prior to selection) were initially screened via PCR and then sequenced to determine which resulted from targeted integration of the donor and to confirm correction of exon 10. Targeted integration was first assessed using one primer that annealed to sequences in intron 9 (but not present within the donor) and one primer that annealed to sequences within the selectable cassette

FIGURE 3.6 Efficacy of ZFN to create a DSB as assessed by Cel-1. By nucleofection of 2 million cells with either 2 μg of each ZFN DNA expression plasmid or 1.5 μg of each ZFN RNA, we are able to achieve ~10% cutting. Uncut bands are indicated by an arrowhead and the two bands expected from digestion of the amplicon by the Surveyor nuclease is indicated by arrows. Percentage of cutting is indicated below each lane. Genomic DNA from non-transfected cells was included in the analysis as a negative control



(primers 1 and 1' in FIGURE 3.7a, FIGURE 3.7b). This resulted in the identification of 21 clones from a total of two experiments (FIGURE 3.7b, TABLE 3). The 21 clones were further characterized by PCR amplification using primers that annealed to the selectable cassette and to sequences downstream of the homology region in intron 10, respectively. This confirmed successful insertion in seven of the 21 clones (primers 2 and 2' in FIGURE 3.7a, FIGURE 3.7c, TABLE 3). Sequencing of the 1 and 1' amplicon verified correction of either the ΔI507 or ΔF508 mutations as well as the presence of silent substitutions introduced in the ZFN-R recognition site and used to create a ClaI site (FIGURE 3.8).

To determine which allele was unmodified and which was targeted, we distinguished the unmodified allele from targeted allele using primers that annealed to sequences completely outside the homology sequences in the donor (FIGURE 3.9a). Primers 3/3' yield a ~1.9 kb amplicon for unmodified alleles and a ~4.5 kb amplicon for correctly targeted alleles. By isolation and sequencing of the specific amplicon corresponding to the unmodified allele (1.9 kb), we were able to determine which allele had been left unmodified and consequently which had been targeted (FIGURE 3.9b). For example, iPSC clones 17-2 and 17-21 (for which sequencing of the unmodified allele revealed ΔI507 genotype) were targeted at the ΔF508 allele. Sequencing of the unmodified allele yielded the ΔI507 mutation in three of 21 clones and the ΔF508 mutation in the other 18 clones (summarized in TABLE 3). Thus targeted allele correction did not appear to occur randomly.

As reported by Bock et al. for other human ESC and iPSC lines (*130*), we found that our uncorrected iPSC (clone 17) expressed low levels of *CFTR* detectable by RT-

FIGURE 3.7 Initial molecular analysis for targeted integration in puromycin-resistant clones. Genomic DNA was isolated from individual puromycin-resistant clones and analyzed initially by PCR to determine targeted integration of the donor. **a)** Schema of the two sets of primers used: 1/1' and 2/2'. In each set, one primer anneals to sequences found in *CFTR* but outside of sequences found within the donor molecule (1 and 2') and the other primer anneals to sequences found within – and specific to – the donor molecule (1' and 2). **b)** Shown are 21 puromycin-resistant clones that yielded an amplicon of the expected size (1.8 kb) with primers 1/1', two clones that did not, the uncorrected iPS 17, MEFs, and a no DNA negative control. **c)** Of the 21 clones that were initially identified with primers 1/1' in panel B, 7 amplified with primer set 2/2'.

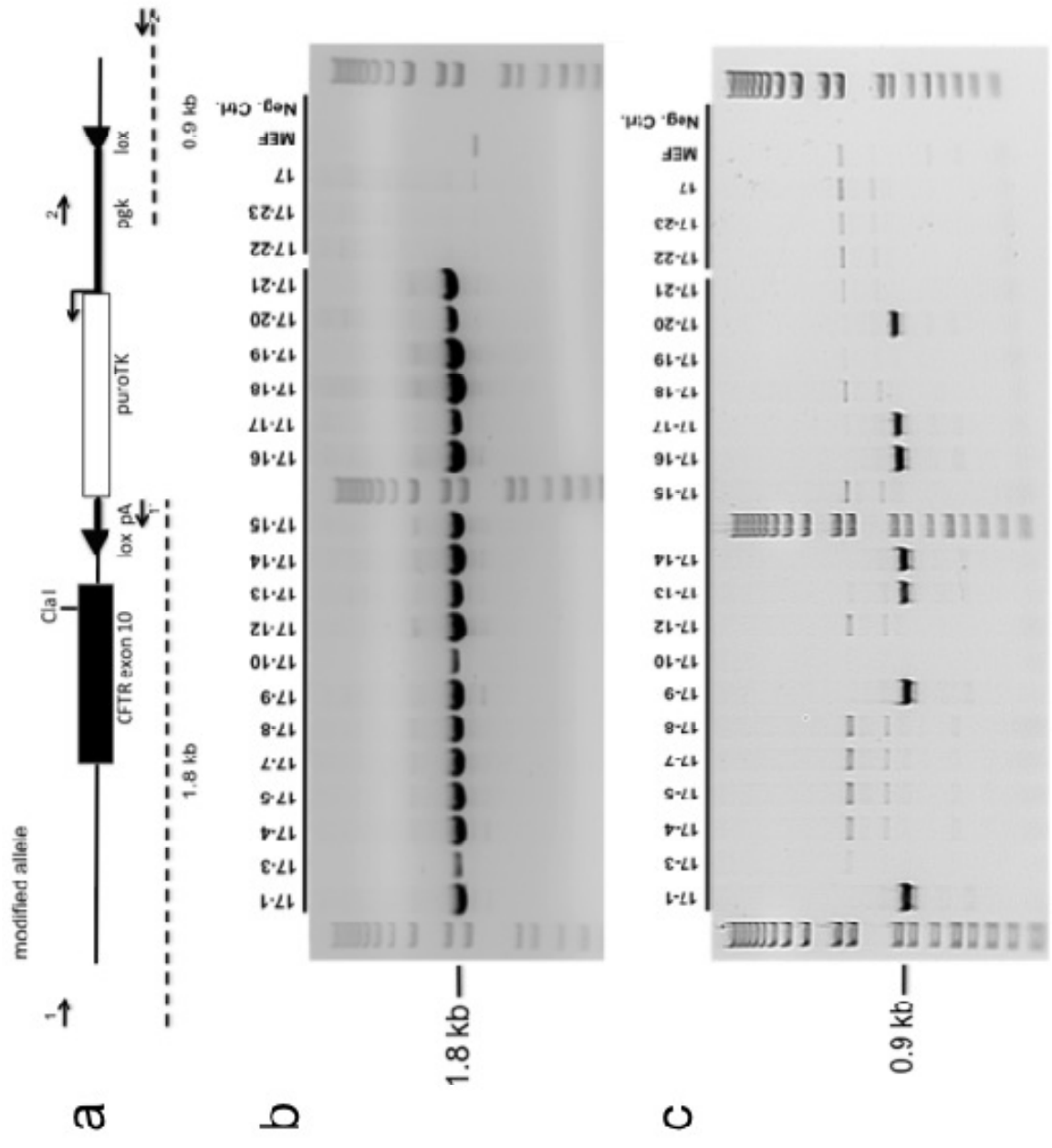

a
modified allele
1
2
ClaI
CFTR exon 10
lox
pA
puroTK
pgk
lox
1.8 kb
0.9 kb
b
Neg. Ctrl.
MEF
17
17-23
17-22
17-21
17-20
17-19
17-18
17-17
17-16
17-15
17-14
17-13
17-12
17-10
17-9
17-8
17-7
17-5
17-4
17-3
17-1
1.8 kb
c
Neg. Ctrl.
MEF
17
17-23
17-22
17-21
17-20
17-19
17-18
17-17
17-16
17-15
17-14
17-13
17-12
17-10
17-9
17-8
17-7
17-5
17-4
17-3
17-1
0.9 kb

TABLE 3 Summary of two experiments to correct CFTR. Between two experiments, a total of 65 puromycin resistant colonies, conferred by integration of the donor, were clonally isolated, expanded, and analyzed for targeted integration of the donor into the CF locus. Indicated are the number of clones that passed each criteria and in brackets are the number of those that showed targeting of the ΔF508 allele. By our first criterion, out of a total of 21 that passed, only 3 showed targeting of ΔF508. In the end, we had 4 clones for which we had evidence by PCR, RT-PCR analysis, and sequencing. All four were targeted at the ΔI507 allele.

	Exp. 1	Exp. 2
65 puro^R clones analyzed	16	49
21 (33%) satisfied 1/1' criterion [3 TI in ΔF508, 18 TI ΔI507]	10 [2 ΔF508]	11 [1 ΔF508]
7 (11%) satisfied 1/1' and 2/2' criteria [all TI in ΔI507]	2	5
4 satisfied 1/1', 2/2', and cDNA criteria [all TI in ΔI507]	2	2

FIGURE 3.8 Sequencing of putative targeted clones confirm targeted integration of donor. The donor contained two silent mutations to the right ZFN (ZFN-R) binding sequences (underlined in blue) of the donor: T to A and A to C (arrows). Downstream of the sequences to correct the ΔI507/ΔF508 mutations, a ClaI restriction enzyme site (underlined in red) was introduced by silent mutation of one bp (arrowhead). Sequences of targeted clones show introduction of these three changes, further confirming targeted integration of the donor.

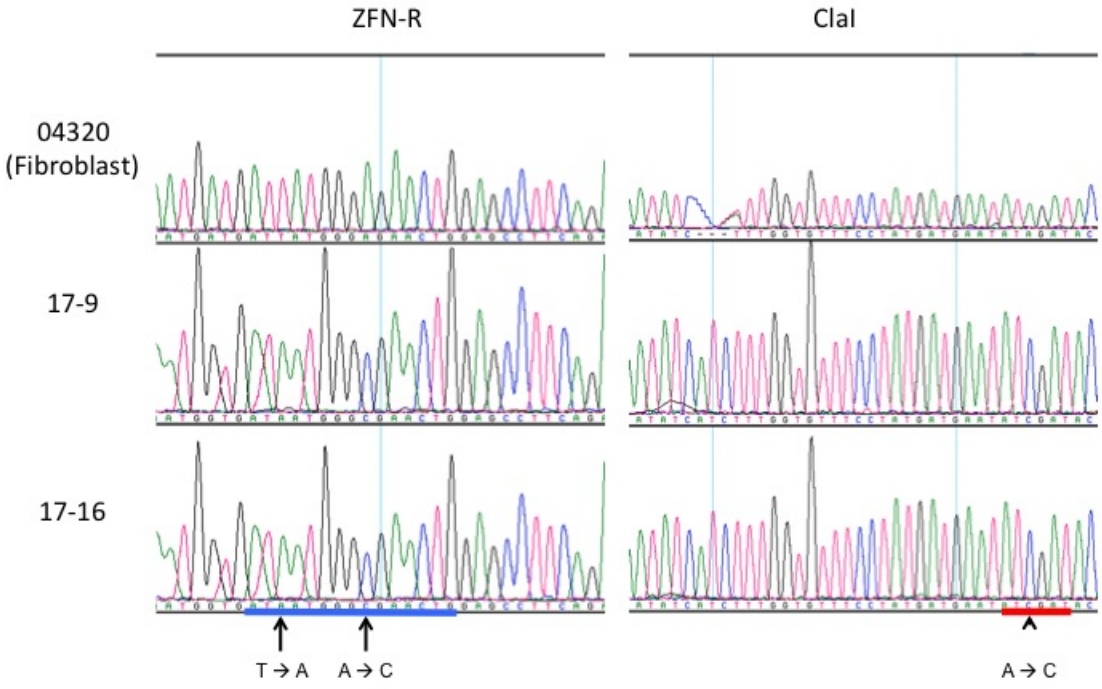
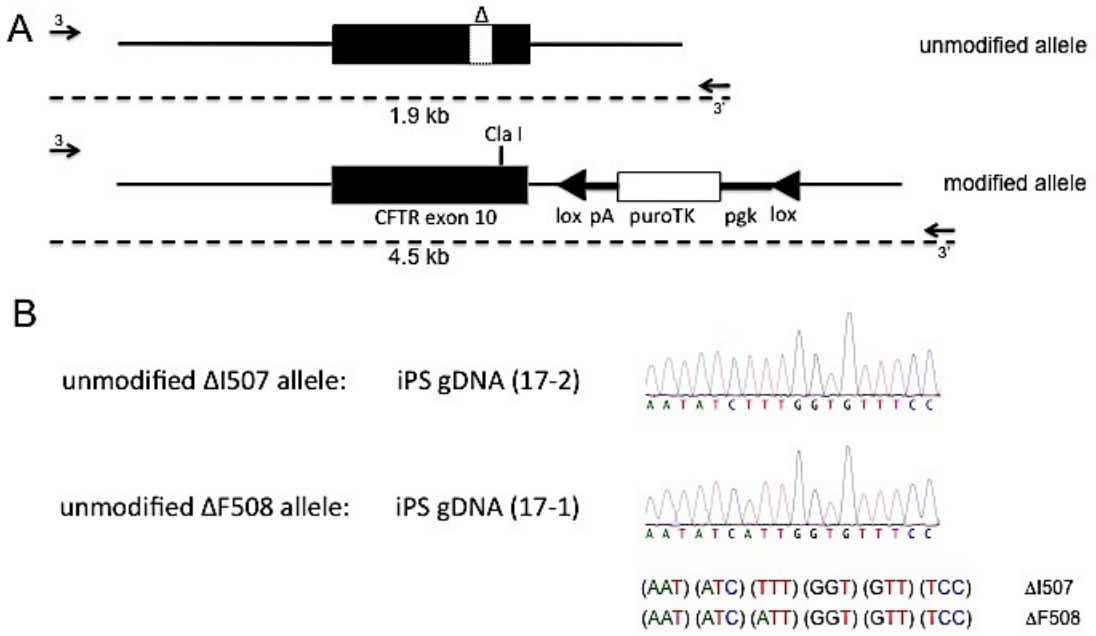


FIGURE 3.9 Sequencing of the unmodified to determine at which allele targeted integration occurred. a) Primers (3/3') that anneal to sequences outside the donor homology sequences were designed to distinguish the unmodified allele from the modified allele. By PCR amplification with these primers, the unmodified allele would yield a 1.9 kb amplicon and the modified allele a 3.5 kb amplicon. **b)** Isolation and subsequent sequencing of the unmodified allele enabled us to determine for each clone in which allele targeted integration occurred. Indicated below the chromatograms are the sequences for ΔI507 and ΔF508.



PCR (FIGURE 3.10a). Sequencing demonstrated a mixture of expression arising from both the ΔI507 and ΔF508 alleles. Similarly, four (17-1, -9, -14, and -16) of the seven clones that passed previous PCR tests for targeted integration (FIGURE 3.7c, TABLE 3) yielded a single band of the same size (0.77 kb) as that seen for the original, uncorrected clone 17 and human ES WA09 cell lines. Sequencing of the amplicons confirmed the expected cDNA organization as well as expression from both unmodified and targeted alleles (FIGURE 3.10b and c). The other three clones (17-13, -17, and -20) also yielded an additional RT-PCR band of higher molecular weight, sequencing of which revealed abnormal cDNA organization for the targeted allele; for these clones there were two copies of exon 10 – either wt exon 10 together with wt exon 10 or wt exon 10 together with ΔI507 exon 10 (e.g. shown for 17-13 in FIGURE 3.11). We did not explore any further to determine how these events might have occurred, but these three clones were no longer considered. Table 3 summarizes the results for each PCR and RT-PCR criteria from two independently performed experiments. All four successfully targeted clones (17-1, -9, -14, and -16) corrected the ΔI507 mutant allele to wt, leaving the ΔF508 allele unmodified (FIGURE 3.12). Therefore, despite gene correction, we did not accomplish our original goal of correcting mutant ΔF508.

Using a probe for the pgk-puro-TK selectable cassette, Southern blot analysis was performed on genomic DNA isolated from the successfully targeted clones (as determined by having passed all criteria listed in TABLE 3; 17-1, -9, -14, -16) to determine whether any showed off-target integration of donor sequences (FIGURE 3.13). The four successfully targeted clones showed no off-target integration. Clones 17-17 and

17-20, which were previously eliminated based on abnormal *CFTR* mRNA patterning, **FIGURE 3.10 Correct RNA expression of CFTR in targeted clones. a)** RT-PCR analysis of *CFTR* mRNA expression in the seven clones that showed targeted integration based on the two PCR analyses (cf. Figure 3.7 and Table 3). Also shown are *CFTR* expression for the original uncorrected iPSCs (17), iMEF, and WA09 cells. Of the seven clones, four (17-1, -9, -14, and -16) show an RT-PC product of the expected size (769 bp) while the other three (17-13, - 17, and -20) also show expression of an additional larger cDNA. **b)** Sequencing of amplified cDNA reveals a mixture of ΔI507/ΔF508 from the original, uncorrected iPS 17 and a mixture of wt/ΔF508 from those clones showing correct RNA expression, here clone 17-1. **c)** Schema of the amplicon from corrected and uncorrected ΔF508 alleles arising from those clones that showed the appropriate *CFTR* mRNA expression. Arrows indicate location of primers used for amplification.

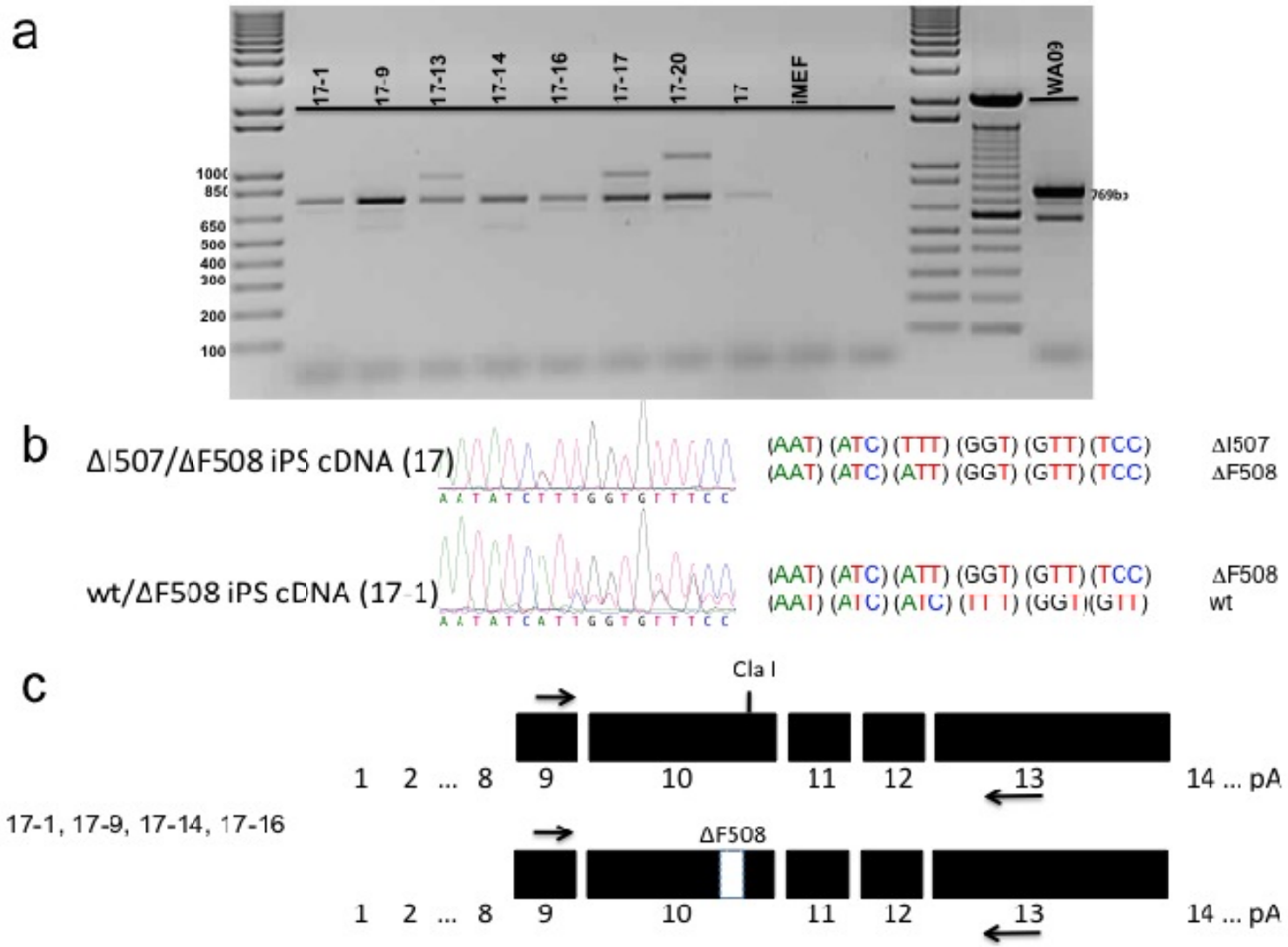

a
17-1
17-9
17-13
17-14
17-16
17-17
17-20
17
iMEF
WA09
1000
850
650
500
400
300
200
100
b
Δ1507/ΔF508 iPS cDNA (17)
(AAT) (ATC) (TTT) (GGT) (GTT) (TCC) Δ1507
(AAT) (ATC) (ATT) (GGT) (GTT) (TCC) ΔF508
wt/ΔF508 iPS cDNA (17-1)
(AAT) (ATC) (ATT) (GGT) (GTT) (TCC) ΔF508
(AAT) (ATC) (ATC) (TTT) (GGT) (GTT) wt
c
Cla I
1 2 ... 8 9 10 11 12 13 14 ... pA
17-1, 17-9, 17-14, 17-16
ΔF508
1 2 ... 8 9 10 11 12 13 14 ... pA

FIGURE 3.11 Incorrect mRNA expression of CFTR in some targeted clones. Both the lower band (blue box a for clones 17-13, 3, 4, 5, and 7)) and the upper band (blue box b for clones 17-13, 3, and 4) were isolated and sequenced. The sequence of the lower band reflected uncorrected mRNA (in these cases, ΔF508). The upper bands, on the other hand, reveal two copies of exon 10: either two wild-type exon 10 (17-4b) or a mixture of wild-type and mutant ΔI507 (17-13b and 17-3b). Arrows indicate PCR primer locations.

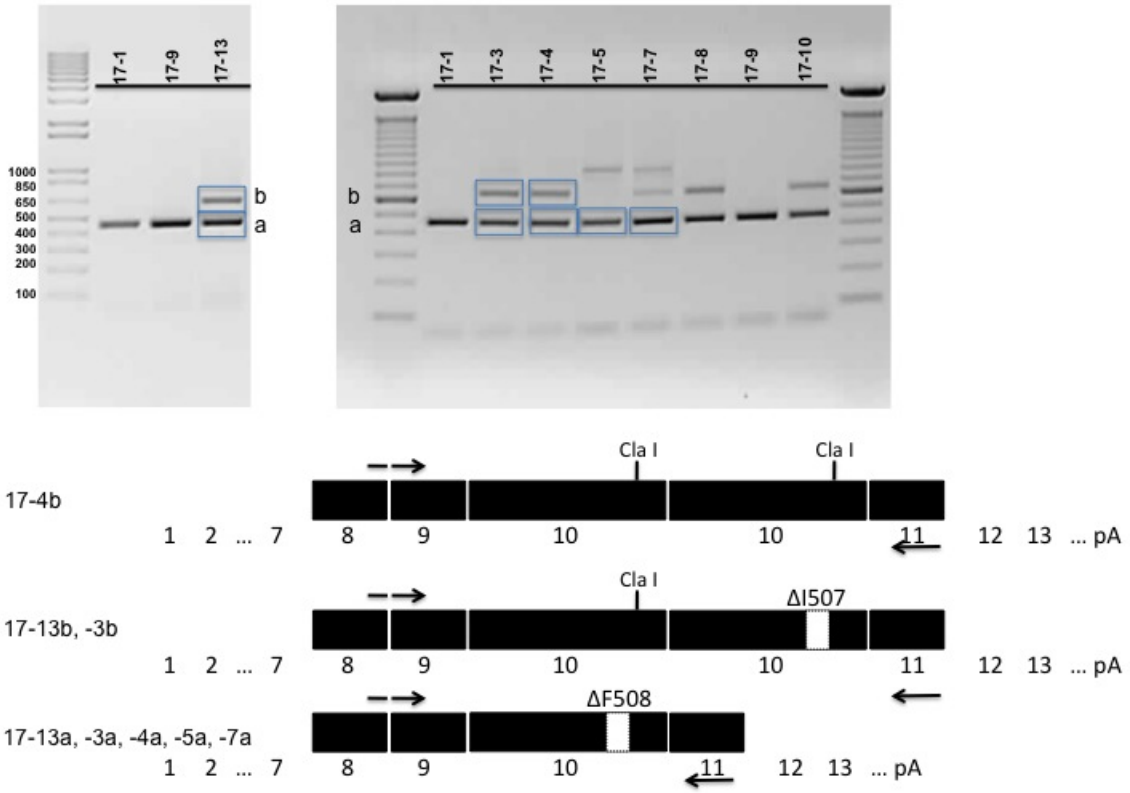


FIGURE 3.12 Sequencing of 4 correctly targeted iPSCs reveal wt/ΔF508 genotype. Isolation of the modified allele (1 and 1'; Figure 3.7) and the unmodified allele (3 and 3'; Figure 3.9) reveals wt/ΔF508 genotype for each of the four corrected iPSC clones showing appropriate RNA expression (Figure 3.10 and Table 3).

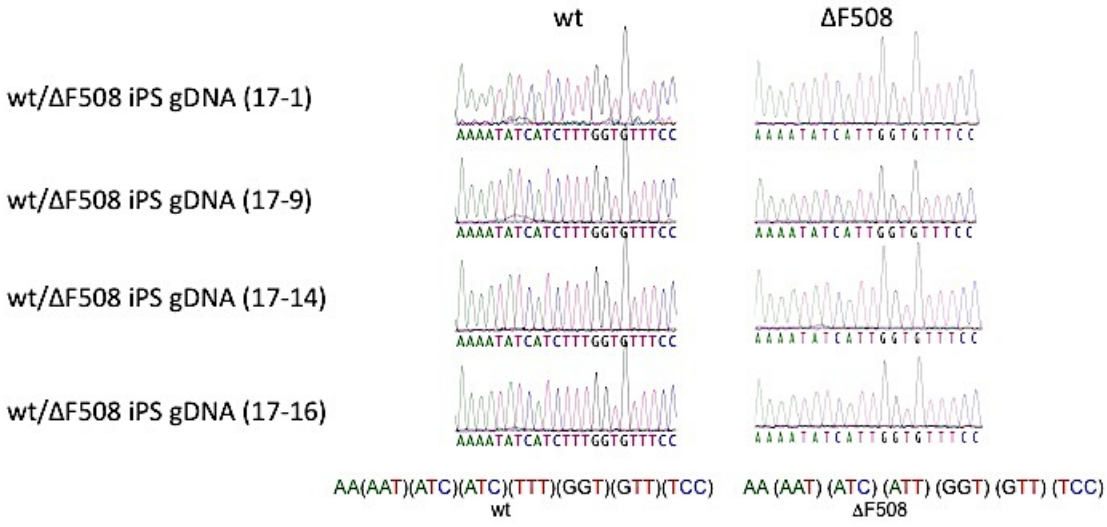
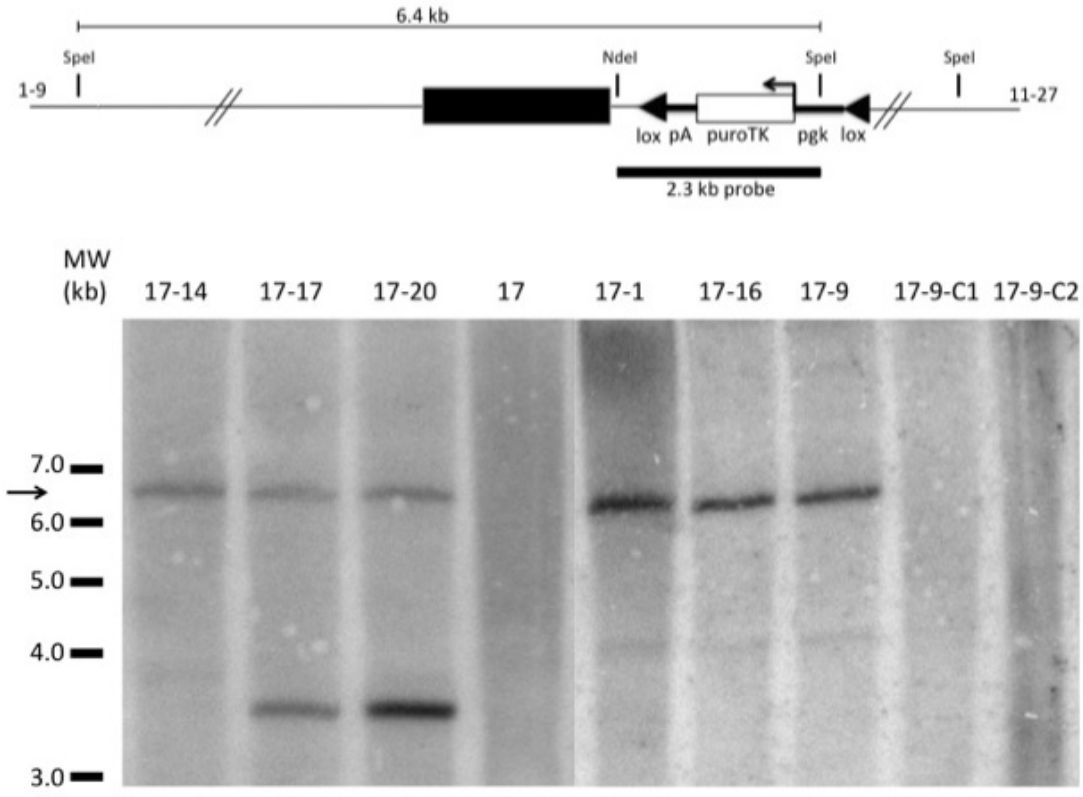


FIGURE 3.13 Southern blot analysis of targeted clones. Above, a schema showing the expected organization of the targeted allele. A 2.3 kb probe from NdeI to SpeI was used to detect the pgk-puro-TK selection cassette. When genomic DNA was digested with SpeI, the probe would hybridize with a 6.4 kb band if accurate targeting occurred. Below, Southern blot results indicate 4 correctly modified clones (17-4, 17-1, 17-16, and 17-9) and two incorrectly modified clones (17-17 and 17-20). In addition to the 6.4 hybridizing band, a band of approximately 3.5 kb was present for these latter two clones, both of which also previously showed incorrect mRNA organization. As expected, the probe did not hybridize with genomic DNA from the original, uncorrected iPS 17. Again, as expected, the 6.4 hybridizing band is also absent in two Cre-excised clones (17-9-C1 and 17-9-C2).



showed additional integration of pgk-puro-TK sequences. Though no attempts were made to elucidate the location of the additional integration, we believe that this additional integration of pgk-puro-TK sequences is also likely within the targeted CFTR locus, consistent with the detection of additional copies of exon 10 in the *CFTR* cDNA from these clones (i.e. the additional RT-PCR bands in FIGURE 3.10).

Because the drug selection cassette is lox-P flanked, this enabled us to quickly and easily remove the whole cassette by transient expression of a Cre-recombinase expression plasmid, which was also delivered via nucleofection. Removal of the pgk-puro-TK cassette yielded cells that are resistant to FIAU but not to puromycin. This was further confirmed by PCR analysis using primers (4 and 4') that recognized sequences outside the homology regions (FIGURE 3.14a and b). Using these primers yields the following amplicons: a 4.4 kb amplicon from the modified allele, a 1.8 kb amplicon from

the modified allele after Cre-excision, and a 1.8 kb amplicon from the unmodified allele. The modified allele after Cre-excision and the unmodified allele can be distinguished by digestion with ClaI, which cleaves modified alleles twice (but not unmodified alleles). RT-PCR analysis of undifferentiated, corrected iPSCs confirmed appropriately sized *CFTR* cDNA, sequencing of which showed expression of CFTR mRNA from both wt and ΔF508 alleles (FIGURE 3.14c and d).

We then tested whether the gene editing and/or Cre-excision methodologies adversely affected the pluripotency or ESC-like transcriptome profile of the iPSCs. We demonstrated that two of the corrected, Cre-excised iPSC clones (17-9-C1 and 17-14-C1) were still pluripotent, able to generate teratomas consisting of all three germ layers, and maintained normal karyotype (FIGURE 3.15). We also observed similar levels of total

FIGURE 3.14 Targeted integration of donor and subsequent Cre-mediated excision of selection marker to correct CF-iPSCs. a) Schema of expected PCR amplicon sizes when using primers (4/4') that anneal to *CFTR* sequences outside of the donor homology arms. Integration of the donor will not only introduce corrective sequences and the selection cassette but also a ClaI site in exon 10. Cre-mediated excision of the selection cassette will leave one loxP site in intron 10. **b)** The upper panel shows the amplicons for the original, uncorrected iPS 17, two corrected clones (17-9 and 17-16), and two Cre-excised clones for each corrected clones (17-9-C1 and 17-9-C2; 17-16-C1 and 17-16-C2). Excision of the selection cassette is confirmed by the presence of just the 1.8 kb band, whereas before excision there are two bands. This is further confirmed in the lower panel, which shows these amplicons digested by ClaI. As indicated in the schema above, ClaI would cut a modified allele two times if the selection cassette is still present but only

once if the cassette has been excised and not at all if it the allele has not been modified. **c)** RT-PCR analysis confirming correct *CFTR* expression in corrected clones before and after Cre-excision as well as uncorrected iPS 17 and WA09. In addition to the expected band (0.46 kb), WA09 also shows a smaller size band. We did not examine this further. **d)** Sequencing of the *CFTR* RT-PCR product in panel C revealed expected mixture of wt and ΔF508 sequences in corrected clones before (17-9 and 17-16) and after (17-9-C1 and 17-9-C2; and 17-16-C1 and 17-16-C2) Cre-excision.

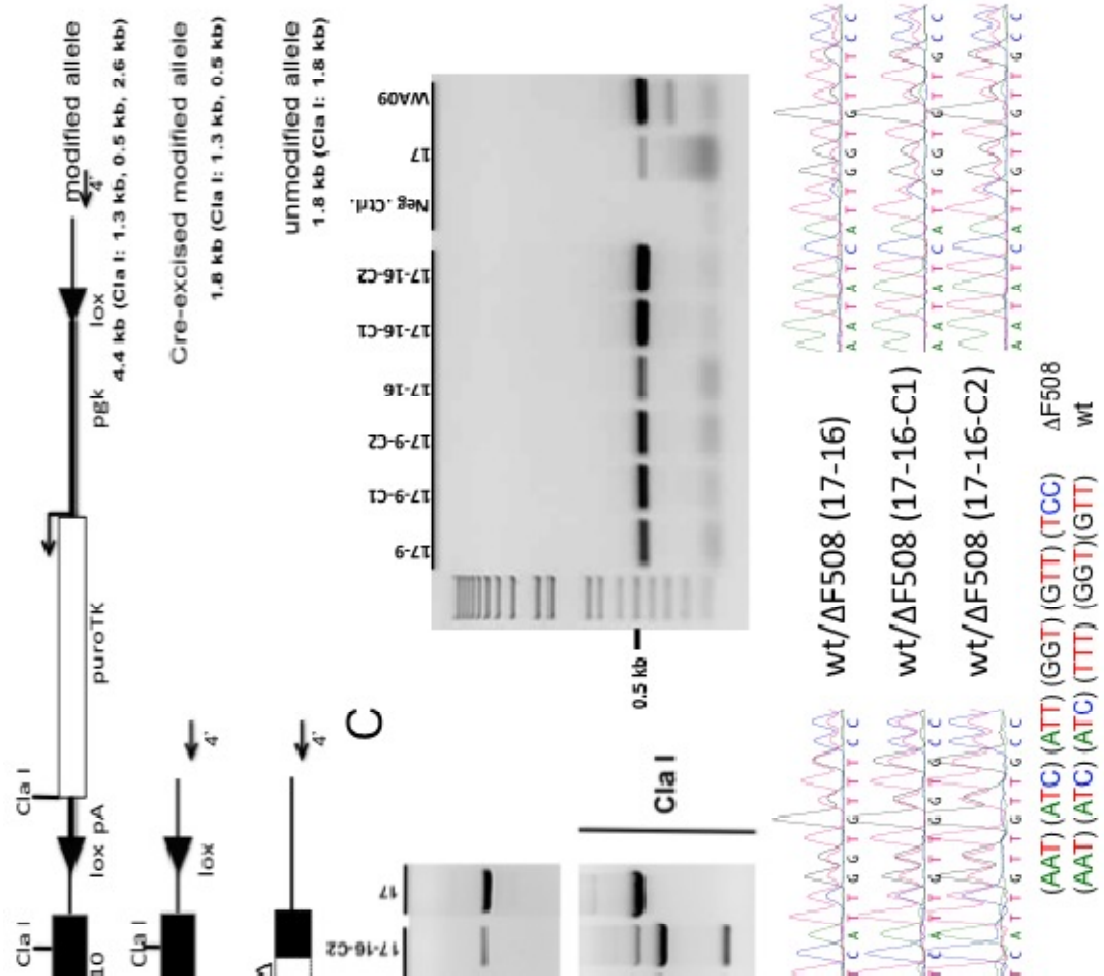
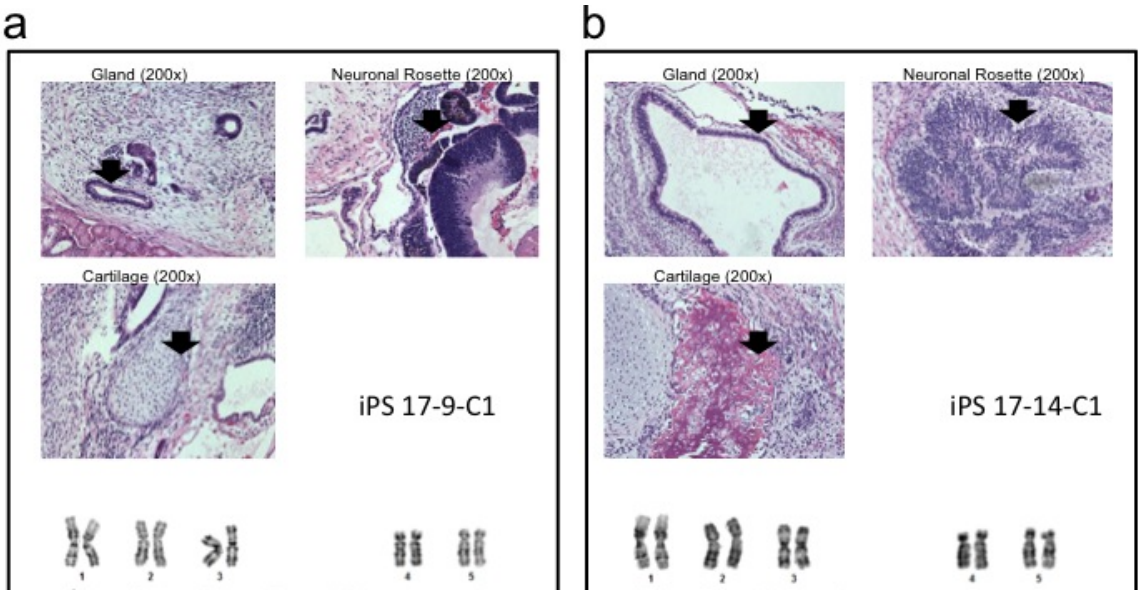


FIGURE 3.15 Gene editing and Cre-excision do not alter pluripotency nor introduce gross chromosomal abnormalities. Above are hematoxylin and eosin staining of teratomas formed by two representative corrected and Cre-excised clones (iPS 17-9-C1, a; iPS 17-14-C1, b). Shown are examples of endoderm (gland), ectoderm (neuronal rosette) and mesoderm (cartilage). Normal male karyotype for each of the two clones is shown below.



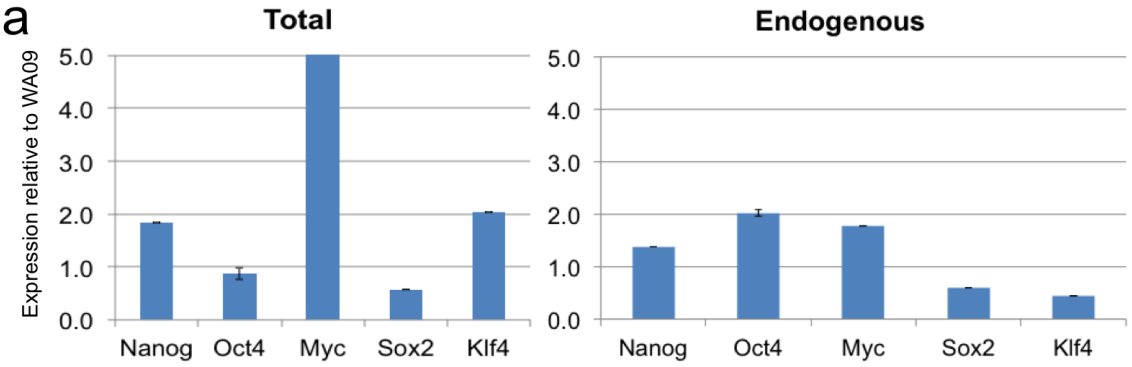
and endogenous mRNA expression of reprogramming factors in Clones 17-9-C1 (FIGURE 3.16a) and 17-14-C1 (FIGURE 3.17a) relative to expression levels in WA09 cells, indicating continued silencing of reprogramming transgenes in these clones. We do note, however, an approximate 5-fold increase in total *C-MYC* expression in 17-9-C1 relative to WA09; a similar increase was not seen for 17-14-C1 Additionally, we assayed the expression of 44 genes characteristic of human pluripotent stem cells and found similar expression patterns between the two corrected clones and the original, unmodified CF-iPS 17 (FIGURE 3.16b and FIGURE 3.17b); in addition, the transcription profiles of the corrected clones remained similar to that of WA09 (FIGURE 3.17c and FIGURE 3.17c).

To evaluate genomic integrity after reprogramming and correction, we first confirmed maintenance of normal karyotype for the two corrected, Cre-excised clones

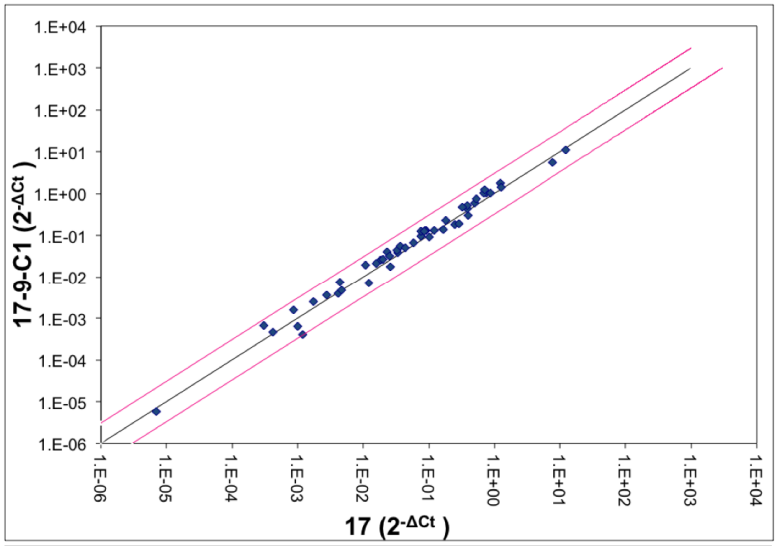
(FIGURE 3.15). For a more detailed analysis, we then submitted genomic DNA from the original CF fibroblast cells, mutant CF-iPSCs (17), and two corrected and Cre-excised iPSC clones (17-9-C1 and 17-14-C1) for complete genome and whole exome sequencing. Both platforms confirmed correction of the ΔI507 (but not the ΔF508) allele in both corrected iPSC clones, while fibroblasts and uncorrected iPSCs had both mutations. We also detected in the modified allele the three silent bp substitutions introduced through HDR with the donor.

Using methods derived from the Complete Genomics structural variation/junction detection pipeline, retroviral vector insertion sites were identified. Table 4 summarizes the identified sites, each found in all three iPS lines (uncorrected iPS 17, corrected 17-9-C1 and corrected 17-14-C1): the approximate location, the orientation of insertion, and **FIGURE 3.16 Transcriptional profile of corrected, Cre-excised Clone 17-9-C1. a)** Total (left) and endogenous (right) expression levels of each reprogramming factor was normalized to expression level of GAPDH and compared to WA09. Comparison of expression levels of 44 genes characteristic of human ESCs in our corrected, Cre-excised clones vs either uncorrected Clone 17 iPSCs **(b)** or WA09 cells **(c)**. The red lines indicate 3-fold changes in gene expression and the black line indicates a 1-fold change.

a


Total
Endogenous
Expression relative to WA09
5.0
4.0
3.0
2.0
1.0
0.0
Nanog
Oct4
Myc
Sox2
Klf4

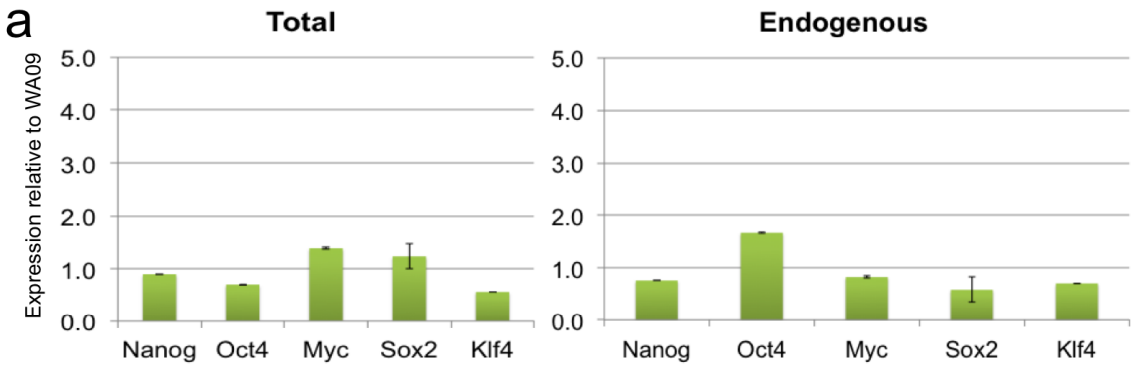
b

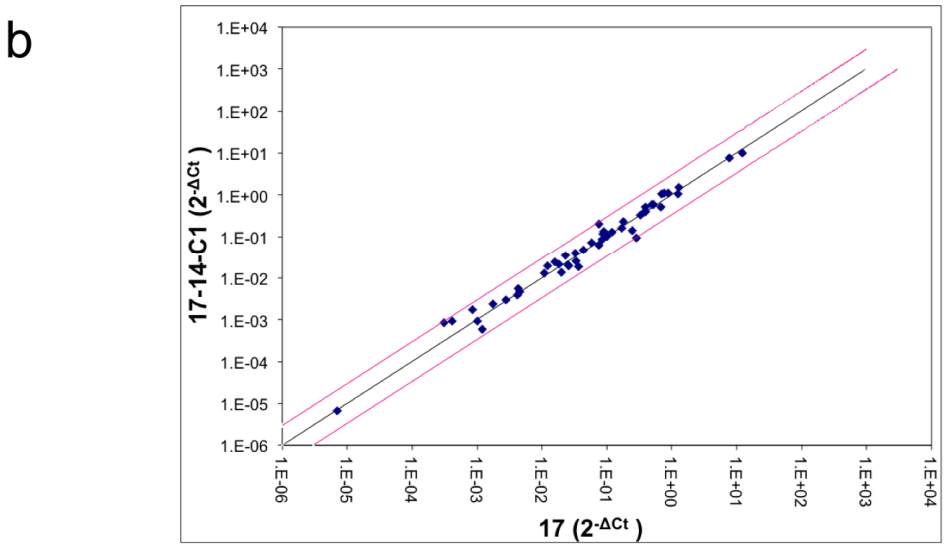

17-9-C1 (2-ΔCt)
17 (2-ΔCt)
1.E+04
1.E+03
1.E+02
1.E+01
1.E+00
1.E-01
1.E-02
1.E-03
1.E-04
1.E-05
1.E-06

c


1.E+04
1.E+03

FIGURE 3.17 Transcriptional profile of corrected, Cre-excised Clone 17-14-C1





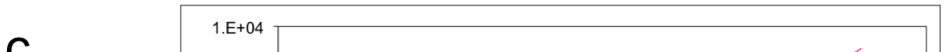


TABLE 4. Viral integration sites identified by complete genome and whole exome sequencing. Twenty-five retroviral integration sites were initially identified by methods derived from Complete Genomics. Indicated are the locations, orientation of integration, and possible genes, mRNA, or EST (as identified by UCSC genome browser) that overlap or are near the integration site. Looking at these specific sites, we found that three (highlighted in orange) are confirmable by our exome sequencing (ExomeSeq). Of these three, based on RefSeq, one site lies within the UTR region of an mRNA, another within intronic region for *EPPIN-WFDC6*, and the third is not found near a gene at all. Chr: Chromosomal location; Left: Left boundary of integration site; Right: Right boundary.

Chr	Left	Right	Direction	Annotation and curation (UCSC genome browser)	Exome
chr2	20343768	20343763	rev	Within mRNA DL492480 (intronic)	Only iPS17 had single read with insertion
chr2	136410548	136410547	fwd	Within R3HDM1 gene, intronic	No coverage
chr2	201672906	?	fwd	Within mRNA AY245869 (intronic)	No coverage
chr2	220271040	220271039	rev	No spanning RefSeq, mRNA per UCSC browser; within H3K27Ac mark (DNaseI hypersensitivity)	No coverage
chr2	242254028	242254025	fwd	Within HDLBP gene, intronic; just upstream (<1kb) of SEPT2; within H3K27Ac mark	No coverage
chr3	48161091	48161086	rev	No spanning RefSeq, mRNA; DNaseI sensitivity in various cells	No coverage
chr3	48755019	48755015	rev	Within mRNA AK097016 (intronic); just upstream (<1kb) of IP6K2; in POL2 ChIPseq signal	No coverage
chr3	48886073	48886068	fwd	Within mRNA AK127190 (intronic); just upstream (<1kb) of PRKAR2A gene	No coverage
chr3	?	48955749	fwd	Insertion into coding sequence, towards end of final coding exon of C3orf71 (<15 aa from stop)	iPS 6/50 Cre9 4/40 14-22 3/30
chr3	49171300	?	rev	Within mRNA HQ447336 (among others); just upstream (<1kb) of LAMB2; SP1 ChIPseq signal; moderate H3K27Ac marking	No coverage
chr3	57961064	57961060	rev	No spanning gene, mRNA, EST; <1kb from end of EST BF509736; within H3K27Ac mark	No coverage
chr3	124932527	124932523	fwd	Within SLC12A8 gene (intronic)	No coverage
chr3	124933484	124933480	rev	Within SLC12A8 gene (intronic, same intron)	No coverage
chr3	188179032	188179036	rev	Within LPP gene (intronic)	No coverage
chr3	197244079	197244199	fwd	Within BDH1 gene (intronic); EXACT LOCATION UNCERTAIN DUE TO TANDEM REPEATS IN REFERENCE, but uncertainty does not change annotation, except possibly placing it in DNaseI hypersensitivity region	1 read Cre9 no ins
chr8	67037263	67037259	rev	Within EST DA715157 (among others); 2kb upstream of TRIM55; within H3K27Ac mark	No coverage
chr10	44387000	44386998	fwd	Within noncoding RNA LOC100506835	No coverage
chr10	104360678	104360673	fwd	Within SUFU gene (intronic); within H3K27Ac mark	No coverage
chr10	?	114712717	fwd	Within TCF7L2 gene (intronic); within H3K27Ac mark	No coverage
chr12	47586916	47586911	fwd	Within FAM113B gene (intronic)	No coverage
chr14	50376987	50376982	rev	Within EST DB236908 (among others)	No coverage
chr16	?	2917513	fwd	Within mRNA AF338196 (intronic); within EST BU690011 (exonic)	No coverage
chr19	17446712	17446711	fwd	Within GTPBP3 gene (intronic)	No coverage
chr20	44175819	44175815	fwd	Within SPINLW1 gene (intronic)	iPS 4/30 Cre9 2/25 14-22 4/19
chr22	46411187	46411183	fwd	DNaseI hypersensitivity; within a few kb downstream of LOC730668 (noncoding pseudogene?)	No coverage

the gene affected as determined by UCSC genome browser. The general integration pattern was consistent with that generally seen for retroviral vectors: that is, preferential integration within transcriptionally active regions (note the number of integrations within histone markers or DNaseI hypersensitive sites). We are unable to assess what effect(s) retroviral integration may have on gene expression or protein function; however, the demonstrated silencing of the vector-expressed transgenes may suggest a decreased potential for the vector to transactivate neighboring genes.

To assess whether the ZFNs had any off-target effects, we looked at the DNA sequences flanking identified novel non-synonymous coding variants (NSCV) for similarity to any permutation of the ZFN target site (TABLE 5). Requiring concordance between whole genome and whole exome sequencing data, we identified one NSCV in the CF-iPSCs, two in 17-9-C1, and seven in 17-14-C1. Importantly, none shared any significant homology to the ZFN recognition sequences, indicating that the NSCVs arose *de novo* rather than as a result of off-target ZFN cutting and subsequent NHEJ. Furthermore, *in silico* detection of off-target ZFN recognition sites by ZFN-Site Search Engine with the implemented FetchGWI source code (*129*), revealed no sequence within the genome closely related to that recognized by the ZFNs (allowing for two bp mismatches per ZFN, a spacer region between 5 and 16 bp, and any combination of ZFN dimerization). In fact, the number of NSCVs observed in the original and corrected iPSCs is completely consistent with that expected from reprogramming and/or long-term culture. Collectively, these evidence support that the CFTR ZFNs we have used are indeed precise, with no off-target cutting observed in coding sequences.

TABLE 5. Non-Synonymous Coding Variants (NSCVs) identified by whole genome and exome sequencing show little homology to CFTR ZFN recognition sequences. NSCVs, almost all of which were single base pair changes, were identified when comparing either corrected iPSC clones 17-9-C1 or 17-14-C1 to uncorrected iPSCs. Shown below the table are the CFTR sequences recognized by the ZFNs (capitalized). Importantly, none of the sequences flanking these NSCVs show any homology to the CF-ZFN target site, suggesting that these NSCVs arose *de novo* independent of ZFN-mediated gene editing.

							WGSeq		ExomeSeq	
Gene_ID	Chr	Position	Variant type	Ref	Mut	Potential ZFN Binding Site	17-9-C1	17-14-C1	17-9-C1	17-14-C1
PRG4 (?)	chr1:186276419-186276459	186276439	snp	G	T	cacccaccaccaccaagtctGcacccaccactaccaaggag				
CYTIP	chr2:158272200-158272240	158272220	snp	C	T	tttcttcctcttccacagcaCgatgaaggccagggataaat				
CLCNKB	chr1:16382249	16382249	ins	-	TG					
SPON1	chr11:14063057-14063097	14063077	snp	G	A	gttttccgtagatcatagacGaagaagaaactcagtttatg				
CLPB	chr11:72141363-72141403	72141383	snp	C	T	cttcttgcatattgttggcaCgggcagcttccaacagggct				
APOBEC1	chr12:7802176-7802216	7802196	snp	G	T	tgtagctaaaaggatgtgtgGcggaatcgtttggtaatggc				
OR6C68	chr12:55886567-55886607	55886587	snp	G	C	aacaaagtgtgcaaaacaatgGttatttgttgttggatggc				
ITGBL1	chr13:102366863-102366903	102366883	snp	G	T	acagagactgcgacaaacatGatggtctcatttgtacaggt				
SHPK	chr17:3527364-3527404	3527384	snp	G	A	tggttgcacagccgaagcccGtggccacactgagatgagac				
ZNF728	chr19:23158575-23158615	23158595	snp	T	A	ttgcaaccttactgaaggctTtgccacattcttcacatttg				
PHACTR3	chr20:58322834-58322874	58322854	snp	C	A	gcaggcaaggccgagaggagCtcatcaagaaggggctgctg				

CFTR CCAGACTTCACTTCTAatggtgATTATGGGAGAACTG

ZFN-L ZFN-R

3.3 Correction of transgene-free CF-iPSCs

Since our initial reprogramming of CF fibroblast cells, there have been significant improvements made in the field, particularly towards making iPSCs that are more therapeutically applicable. As such, we obtained two transgene-free iPSCs (RC202 and RC204) that were originally generated by reprogramming homozygous ΔF508 CF fibroblasts with a lox-P flanked, polycistronic lentiviral vector expressing OCT4, SOX2, KLF4, and c-MYC that was subsequently Cre-excised (*63*). Using the same methodology employed to correct our ΔI507/ΔF508 iPSCs, puromycin resistant transgene-free clones were first assessed by RT-PCR (FIGURE 3.18) and clones showing appropriate *CFTR* mRNA expression were then analysed by PCR to determine targeted integration. As with our CF-iPSCs (FIGURE 3.7a), we observed that some clones that passing the 1 and 1' criterion did not also pass the 2 and 2' criterion (data not shown). In the end, we identified 6 corrected clones that passed all PCR analyses and that were of genotype wt/ΔF508 (FIGURE 3.19).

3.4 Allele-preferred targeting

When we mapped out our method of correcting the ΔI507/ΔF508 CF-iPSCs, we assumed we would observe nearly equal numbers of ΔI507 and ΔF508 corrected clones, since there was no reason *a priori* to expect otherwise. Surprisingly, in two independent experiments (summarized in TABLE 3), we found a significant skewing towards correction of the ΔI507 allele over the ΔF508 allele at a 6:1 targeting ratio, respectively. We looked to see whether there were any differences in the sequences of the two alleles, paying particular attention to the 1.6 kb homology sequences present in the donor. In doing so, we found that there was a single bp difference in intron 9, 76 bp upstream of the

FIGURE 3.18 mRNA analysis of transgene-free iPSCs selected for puromycin resistant colonies. To correct transgene-free iPSCs, the cells were nucleofected to co-deliver ZFNs and donor. Puromycin resistant colonies were clonally isolated and first assessed by RT-PCR to determine appropriate mRNA expression. The clones of interest were those having only a single band of 0.8 kb.

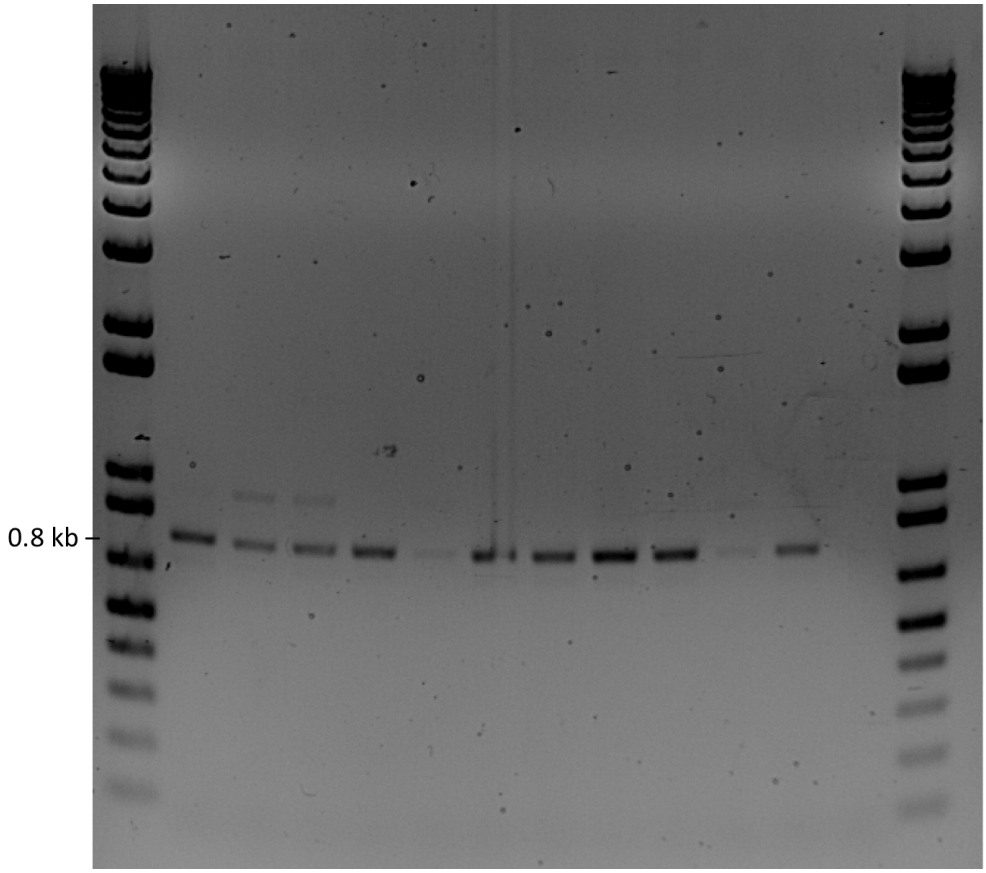
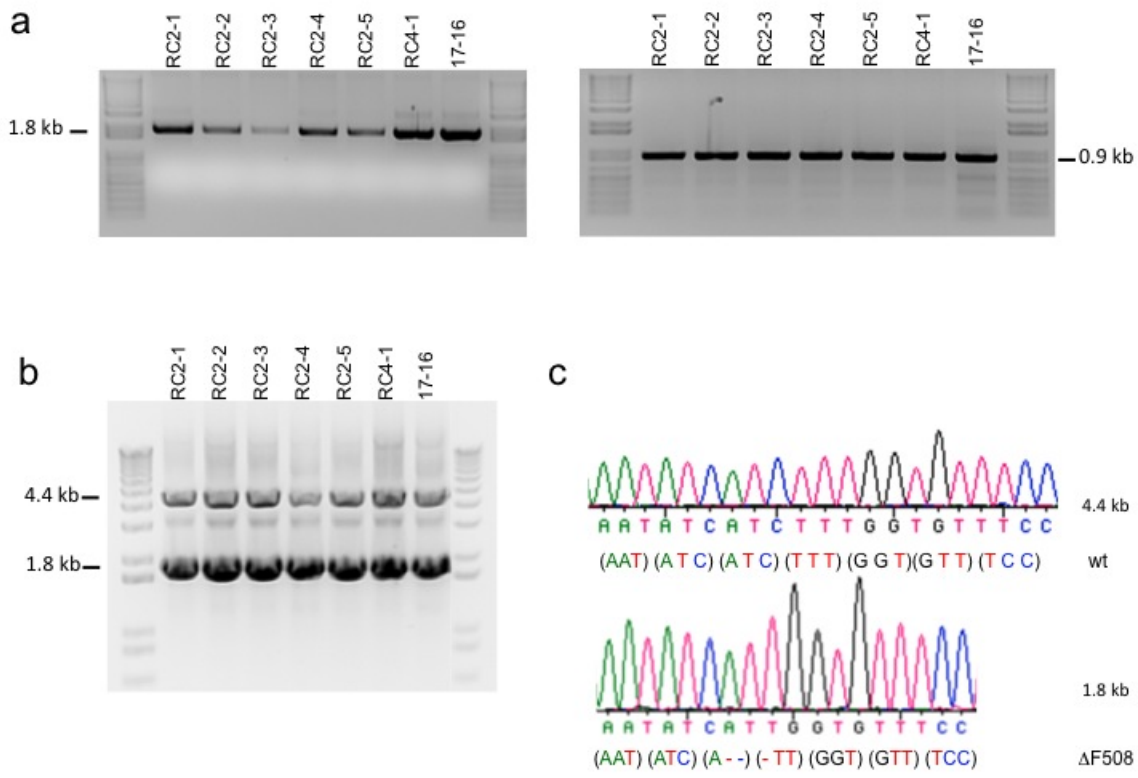


FIGURE 3.19 Correction of transgene-free CF-iPSCs. Correction was accomplished using the same ZFN-mediated HDR approach applied to our ΔI507/ΔF508 iPSCs. **a)** Targeted integration was assessed with the same two sets of primers used in Figure 3.7, wherein one primer of each set annealed *CFTR* sequences outside of the homology arms founds in the donor and the other primer of each set annealed donor-specific sequences. Shown are 6 clones exhibiting targeted integration of the donor in one of two transgene-free iPSC clones (RC202 or RC204) and a ΔI507/ΔF508 corrected clone (17-16). **b)** Using primers annealing *CFTR* sequences outside the homology arms, each sample exhibits two bands: the 4.4 kb band corresponding to the modified allele and the 1.8 kb band to the unmodified allele. **c)** Isolation and sequencing of each band confirms wt/ΔF508 sequencing for each clone (a representative chromatogram is shown).



intended ZFN cleavage site (FIGURE 3.20a). The ΔF508 allele has a guanine, whereas both the ΔI507 and donor have an adenosine at this site. We speculated that if this increased homology between the donor and the ΔI507 allele resulted in increased targeting of the ΔI507 allele over the ΔF508 allele, then a change in the donor so that it matched the ΔF508 allele should now favor targeting of the ΔF508 allele over the ΔI507 allele. Indeed, an A>G substitution in our donor skewed targeting from the ΔI507 allele in favor of ΔF508. Instead of a 6:1 ratio with Donor 1 (the "A donor"), we saw a 1:3.7 ratio for targeting of the ΔI507 allele vs the ΔF508 allele with Donor 2 (the "G donor") (FIGURE 3.20b). This single bp change resulted in a 22-fold net change. As with the original correction of our CF-iPSCs with Donor 1 and correction of transgene-free iPSCs, we observed similar patterns of targeting (ie. some of the clones passed the 1 and 1' criterion did not also pass the 2 and 2' criterion; FIGURE 21a). Using Donor 2 we were able to identify four clones that yielded the appropriately size amplicon in our RT-PCR analysis of the *CFTR* gene (FIGURE 3.21c), two of which showed correction of the ΔF508 allele (FIGURE 3.22, FIGURE 3.23).

3.5 Demonstration of restored protein expression

Based on previous RT-PCR analysis of the *CFTR* gene in undifferentiated corrected iPSCs, we knew that mRNA expression arose from both the unmodified and modified alleles (FIGURE 3.10 and 3.11). We then wanted to explore whether both alleles would be appropriately regulated under *in vitro* differentiation conditions. Based on previous reports that treatment of human ESCs and iPSCs with Activin A induces differentiation into definitive endoderm, from which lung cells are derived during

FIGURE 3.20 Allele-preferred targeting. a) Sequencing of each allele in our ΔI507/ΔF508 fibroblast cells revealed a single base pair difference in the 1.6 kb homology sequences also found in the donor. This SNP is located 76 bases upstream of the intended ZFN cut site. **b)** The donor used in the experiments presented thus far had an A at this position (donor 1), thereby matching the ΔI507 allele more closely. In using this donor, the ΔI507 allele was targeted more than the ΔF508 allele (85.7% vs 14.3%; summarized below the schema of donor 1). In using a donor that had a G at this position and thus matched the ΔF508 allele more closely (donor 2), the ΔF508 allele was now targeted more frequently than the ΔI507 allele.

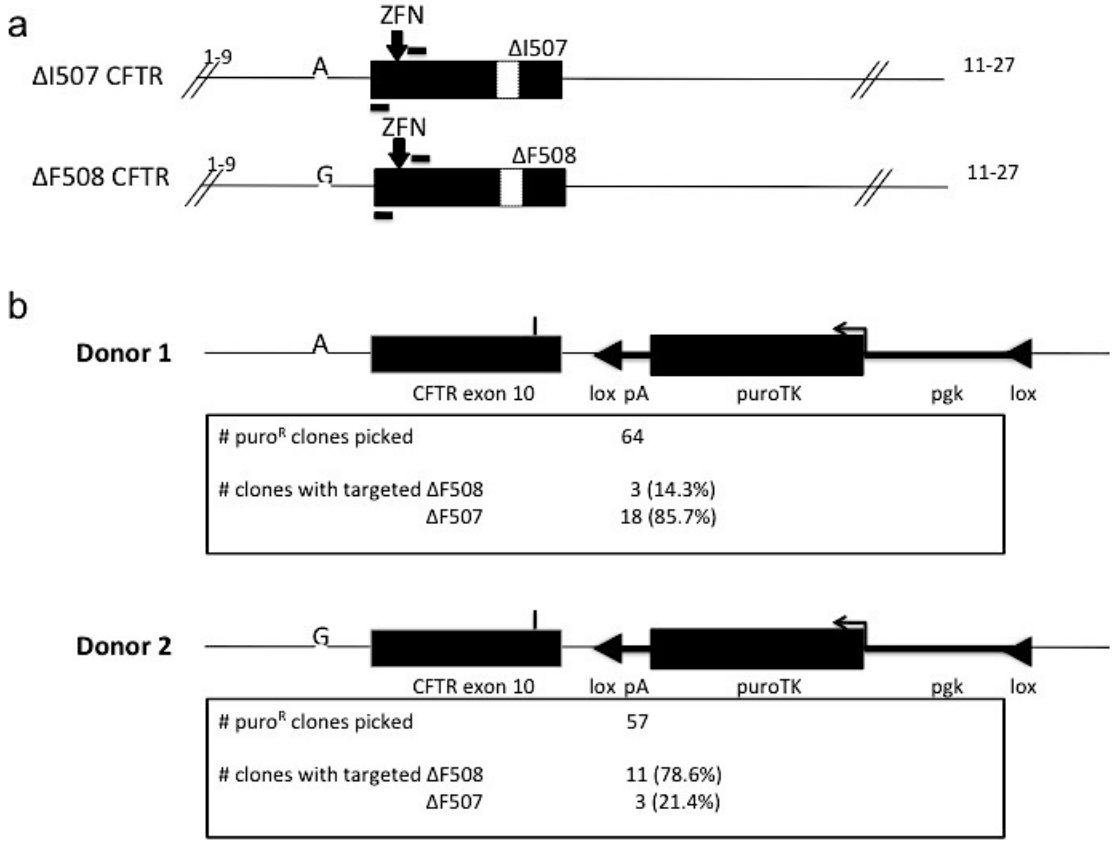


FIGURE 3.21 Targeting of ΔF508 allele with Donor 2. a) Puromycin resistant colonies were assessed for targeted integration first by PCR using 1 and 1' primers (top) and 2 and 2' primers (bottom). **b)** Those that passed both PCR criterion were then assessed for correct *CFTR* mRNA expression by RT-PCR. Four clones showed appropriate mRNA expression.

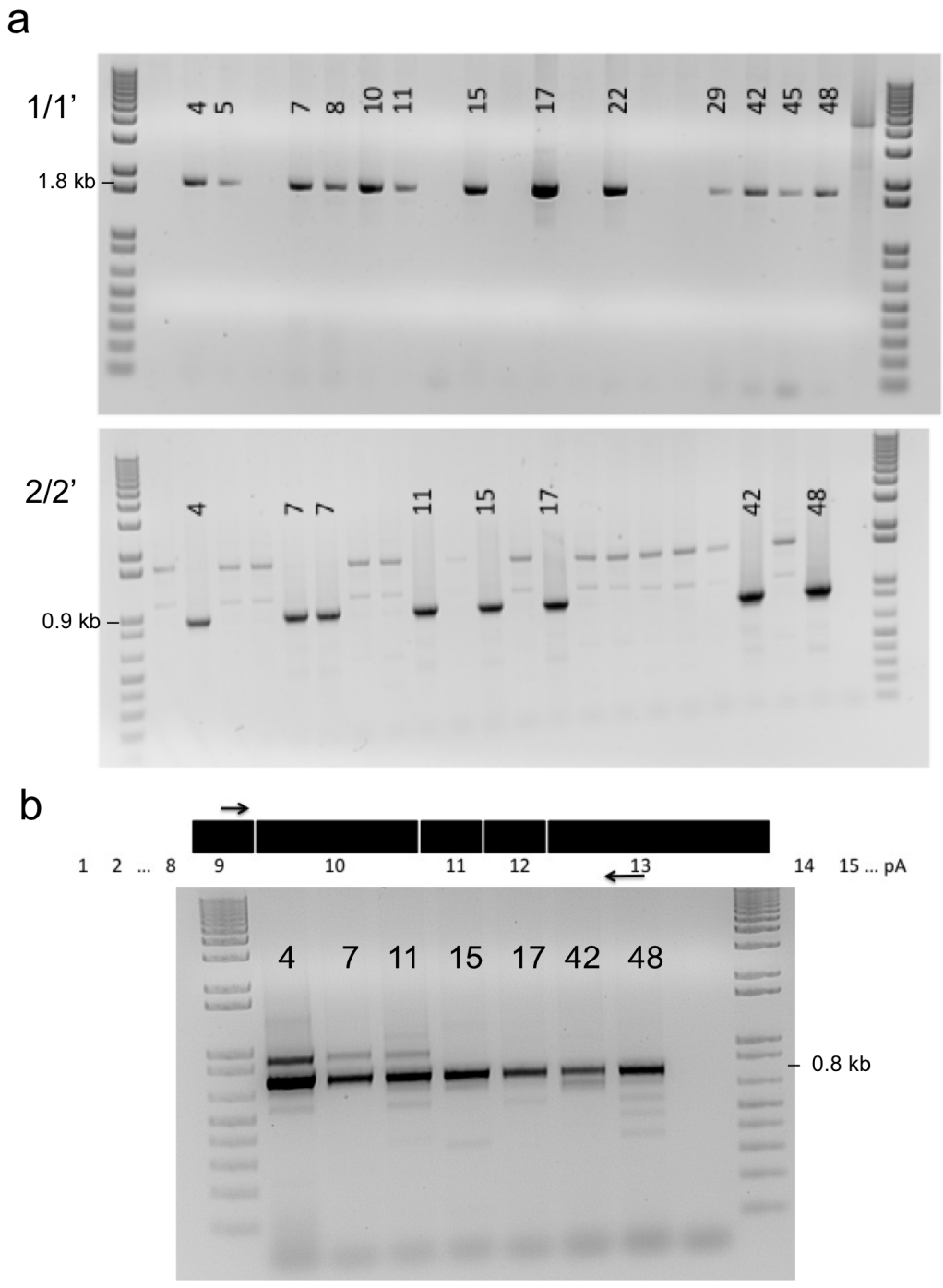


FIGURE 3.22 Correction of ΔF508 with Donor 2. Using the donor matching the ΔF508 allele, ZFN-mediated HDR was employed to correct ΔF508 in our ΔI507/ΔF508 iPSCs. **a)** Integration of the donor as demonstrated by using the same two primers previously used to initially characterize targeted integration (cf Figure3.7). Shown are two representative corrected clones (17-42 and 17-48) and the original, uncorrected iPS 17. **b)** Using the primers that annealed *CFTR* sequences outside the donor homology, the unmodified allele was preferentially amplified, isolated and sequenced. By sequencing, the unmodified allele was identified as ΔI507, indicating that the ΔF508 allele had been corrected.

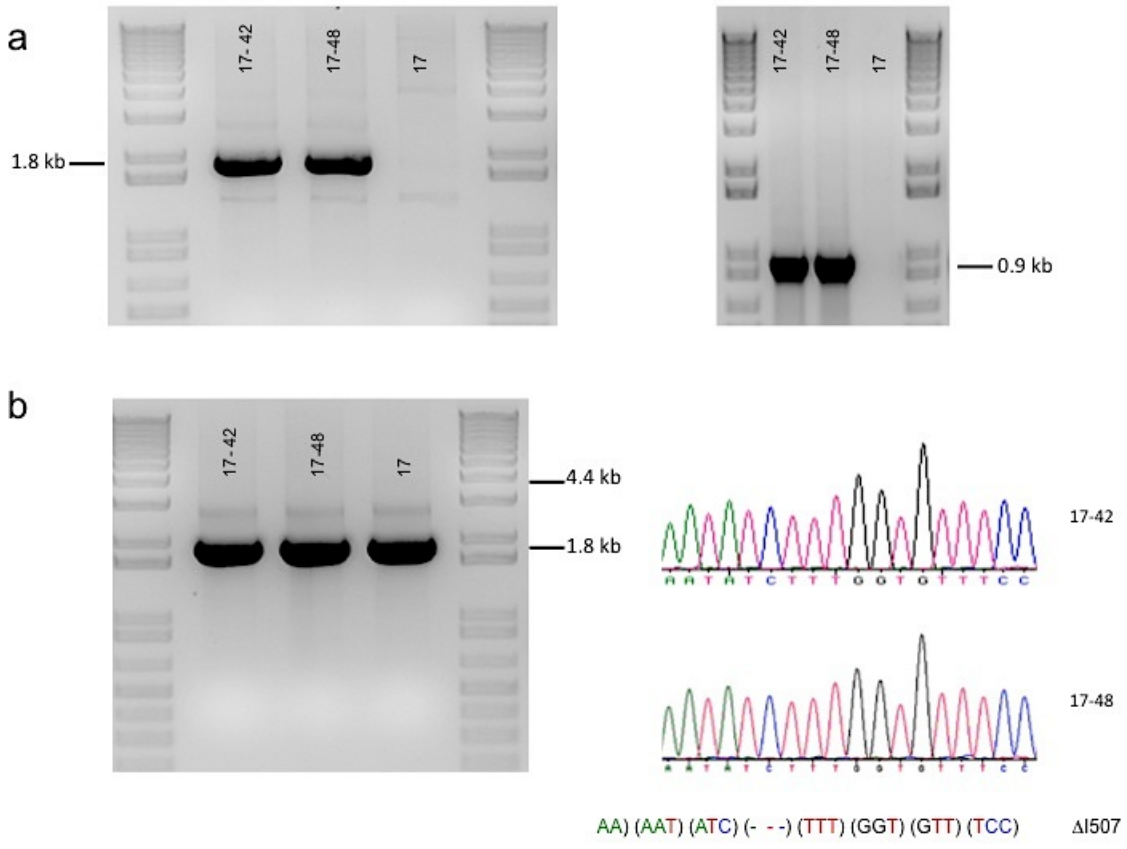
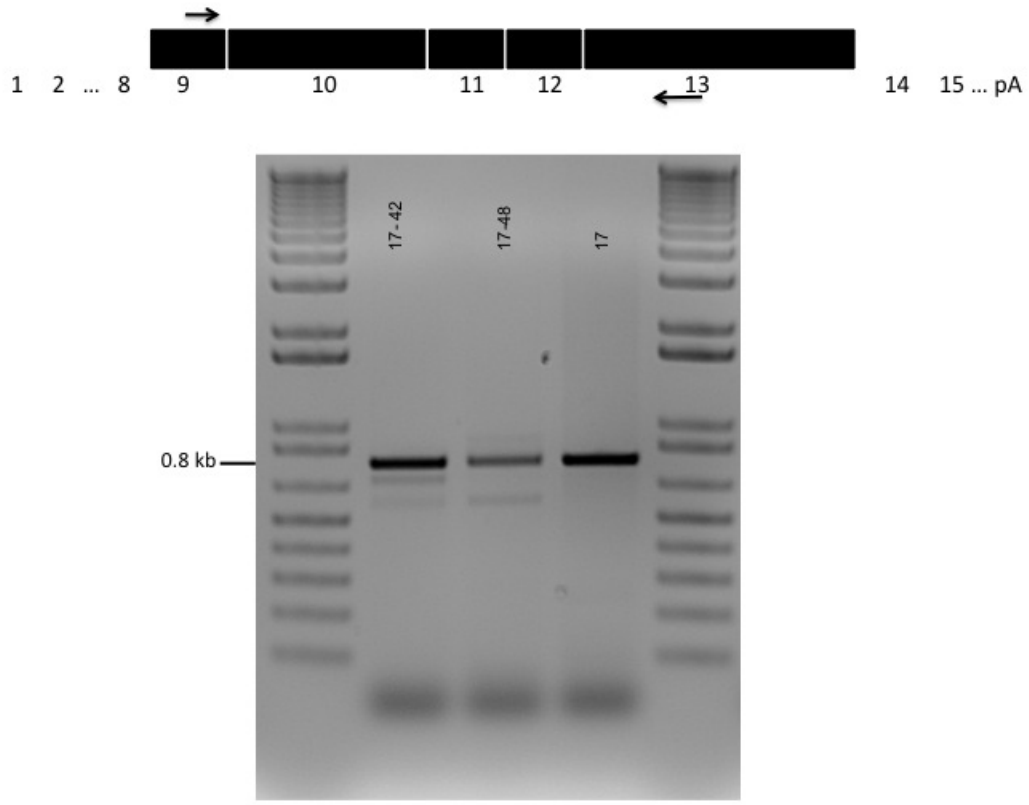


FIGURE 3.23 wt/ΔI507 corrected clones show normal *CFTR* mRNA expression. RT-PCR primers used here anneal to exon 9 and exon 13, resulting in a 0.8 kb band that is exhibited in all samples, both the two corrected clones and the original, uncorrected iPS 17.



embryonic development (*125*, *131*), we treated the original iPS 17 and corrected iPS 17-9-C1 with Activin A for 3-5 days (FIGURE 3.24a, b). In doing so, we saw a moderate up-regulation of *CFTR* and *SOX17* mRNAs, the latter being expressed in definitive endoderm. Sequencing of the RT-PCR *CFTR* amplicon for iPS 17-9-C1 at day 5 of Activin A treatment revealed approximately expression of both the mutant and corrected mRNAs (FIGURE 3.24b). We also explored the use of an air-liquid interface (ALI) culture system that was reported to give rise to tissue that resembled human airway epithelium (*13*, *132*). Under these conditions for corrected clones 17-14 and 17-16, we also saw up-regulation of *CFTR* mRNA (FIGURE 3.24c); sequencing of the RT-PCR amplicon showed expression of both the unmodified (ΔF508) and corrected (wt) mRNAs.

Though the RNA data suggested appropriate transcription of the corrected *CFTR* allele, neither the Activin A induction of definitive endoderm differentiation nor the ALI culture resulted in detectable CFTR protein expression. By adopting a step-wise differentiation of definitive endoderm into NKX2.1-expressing lung and thyroid progenitors as described by Green et al. and Longmire et al. (summarized in FIGURE 3.25a) (*79*, *127*), we were able to differentiate both uncorrected (17) and corrected, Cre-excised (17-16-C1) iPSC clones, as well as WA09 cells, into populations of which 79-95% expressed both CXCR4 and C-KIT, both definitive endodermal markers, after 4 days of differentiation (FIGURE 3.25b).

After subsequent anteriorization and ventralization to derive anterior foregut endodermal cell fates (day 19 in FIGURE 3.25a), we observed increased expression of pan-endodermal marker *FOXA2;* thyroid marker *PAX8;* and lung markers *NKX2*.1, SOX9 (of which coexpression with NKX2.1 indicate distal lung bud epithelium (*133*), and P63

FIGURE 3.24 Corrected *CFTR* mRNA expression by iPSC-derived cells. a) mRNA expression pattern of uncorrected iPS 17 for *CFTR* and *SOX17* over 3 days of Activin-A induction of corrected. **b)** Gene expression pattern of corrected, Cre-excised iPSC Clone 17-9-C1 over 5 days of Activin-A induction show similar upregulation of *CFTR* and *SOX17*. *CFTR* RT-PCR amplicon at day 5 (blue box) was isolated and sequenced confirming co-expression of ΔF508 and wt alleles. **c)** mRNA expression pattern of *CFTR* in uncorrected, undifferentiated (17 iPS) and uncorrected, differentiated (17 ALI) cells as well as two corrected and differentiated cells (17-14 ALI and 17-16 ALI). Differentiation was accomplished by maintaining cells in air-liquid interface culture conditions, whereby clumps of iPSCs were cultured within semi-permeable inserts with ESC media both outside and inside the insert (liquid-liquid conditions) for four days then in differentiation media under liquid-liquid conditions. After 5 days of liquid-liquid differentiation media conditions, the cells were maintained under air-liquid conditions with differentiation media only in the outside the insert. Sequencing of the uncorrected, differentiated and one of the corrected, differentiated samples reveal ΔI507/ΔF508 and wt/ΔF508 sequencing, respectively. GAPDH was used as a loading control.

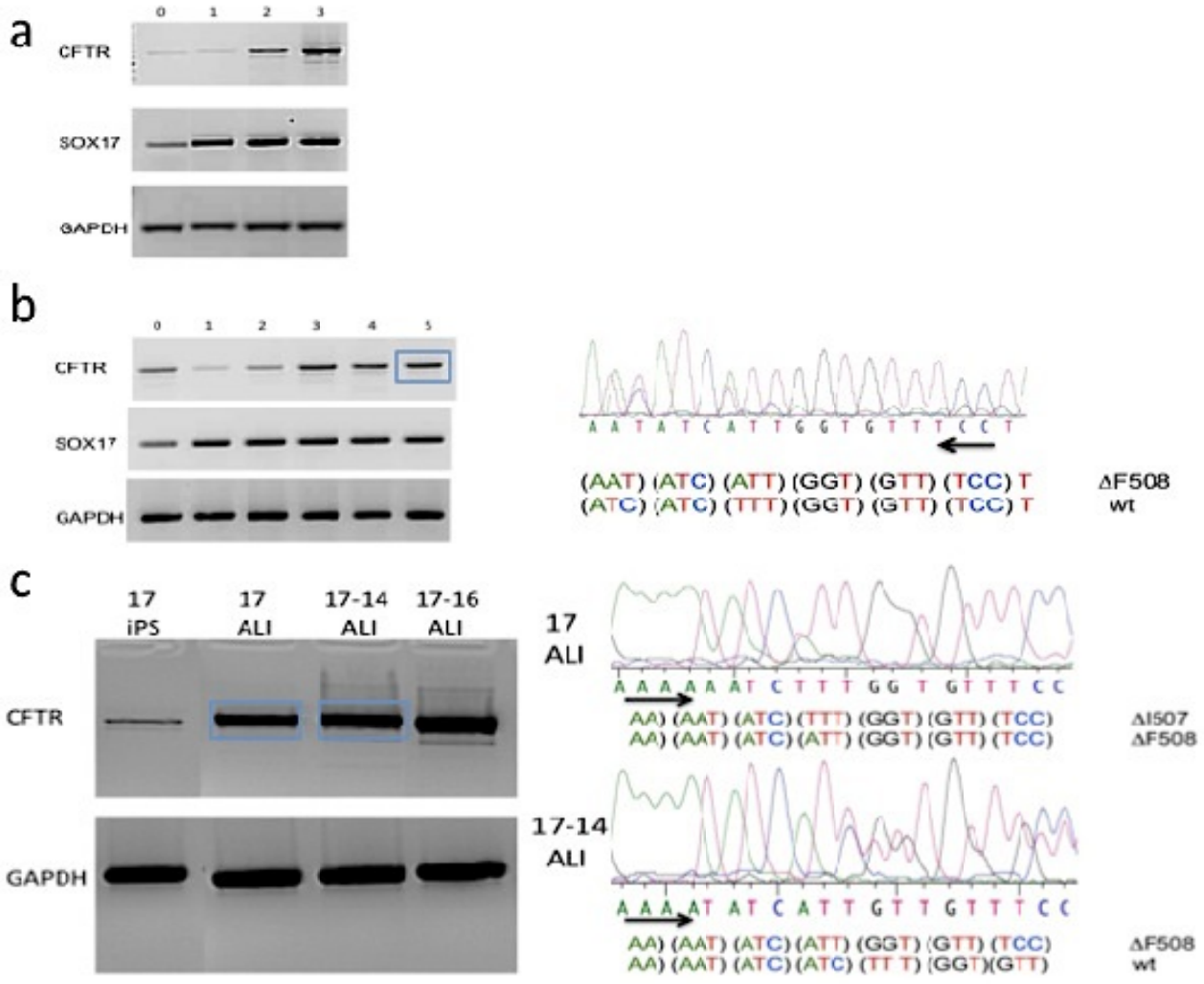
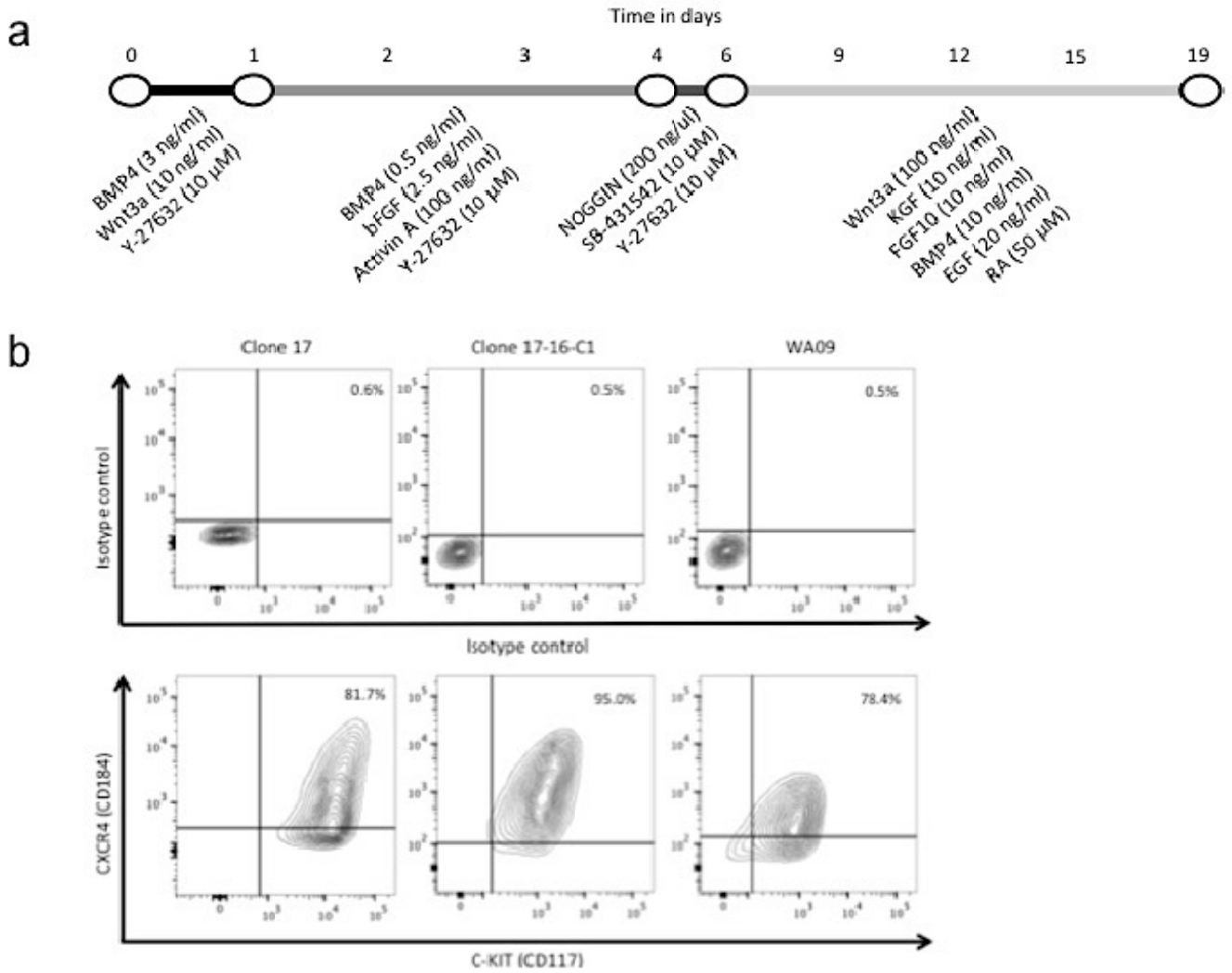

a
0 1 2 3
CFTR
SOX17
GAPDH
b
0 1 2 3 4 5
CFTR
SOX17
GAPDH
(AAT) (ATC) (ATT) (GGT) (GTT) (TCC) T ΔF508
(ATC) (ATC) (TTT) (GGT) (GTT) (TCC) T wt
c
17 iPS
17 ALI
17-14 ALI
17-16 ALI
CFTR
GAPDH
17 ALI
AA) (AAT) (ATC) (TTT) (GGT) (GTT) (TCC) ΔI507
AA) (AAT) (ATC) (ATT) (GGT) (GTT) (TCC) ΔF508
17-14 ALI
AA) (AAT) (ATC) (ATT) (GGT) (GTT) (TCC) ΔF508
AA) (AAT) (ATC) (ATC) (TTT) (GGT) (GTT) wt

FIGURE 3.25 Differentiation of iPSCs into NKX2.1-expressing anterior foregut progenitors to lung and thyroid lineages. a) Summary of the differentiation protocol adapted from Green et al. (2011) and Longmire et al. (2012). Embryoid bodies are formed (Day 0) and differentiated into definitive endoderm, which gives rise to both anterior and posterior foregut, by induction of TGFβ signaling via Activin A treatment (Days 1-4). Then, after a brief inhibition of TGFβ and BMP signaling pathways by SB-431542 and Noggin, respectively, the cells are treated for 11-13 days with a cocktail to ventralize into anterior foregut progenitors that express NKX2.1. **b)** At Day 4 of differentiation, some of the cells are analyzed by flow cytometry for expression of definitive endodermal markers CXCR4 and C-KIT. If the population is ≥ 70% double-positive, we proceed with the anteriorization step. **c)** Expression profile of differentiated cells after 19 days indicate differentiation towards early lung and thyroid endodermal lineages.


a
Time in days
0
1
2
3
4
6
9
12
15
19
BMP4 (3 ng/ml)
Wnt3a (10 ng/ml)
Y-27632 (10 μM)
BMP4 (0.5 ng/ml)
bFGF (2.5 ng/ml)
Activin A (100 ng/ml)
Y-27632 (10 μM)
NOGGIN (200 ng/ul)
SB-431542 (10 μM)
Y-27632 (10 μM)
Wnt3a (100 ng/ml)
KGF (10 ng/ml)
FGF10 (10 ng/ml)
BMP4 (10 ng/ml)
EGF (20 ng/ml)
RA (50 μM)
b
iClone 17
Clone 17-16-C1
WA09
0.6%
0.5%
0.5%
Isotype control
Isotype control
81.7%
95.0%
78.4%
CXCR4 (CD184)
C-KIT (CD117)

(basal cells) (FIGURE 3.26). Significantly, *CFTR* expression increased 10- to 100-fold (depending on the experiment) when compared to undifferentiated cells, though levels of expression in day 19 differentiated cells were still low compared to positive controls (fetal lung, fetal thyroid, and – in the comparison of *CFTR* gene expression – Calu3 cells). Sequencing of the *CFTR* RNA produced by the day 19 differentiated population confirmed the expected CFTR mRNA organization and sequence (i.e. Clone 17 being ΔI507/ΔF508 and Clone 17-16-C1 being wt/ΔF508 while both WA09 and CFTR-expressing Calu-3 cells expressed wt RNA; FIGURE 3.27).

Importantly, we were now also able to detect CFTR protein by western blot in the day 19 cultures (FIGURE 3.28). In CFTR-expressing Calu-3 cells and differentiated WA09, we detected a protein band of ~170 kDa (arrow; - columns) that is consistent with a mature CFTR protein. The mature CFTR protein is sensitive to deglycosylation when treated with an *N*-glycosydase F (arrowhead; + columns), as evident by a reduction in molecular weight to ~140 kDa. We detected only a slight amount of the ~140 kDa protein and none of the 170 kDa protein in cells derived from the uncorrected iPSCs (Clone 17). Significantly, correction of the Clone 17 ΔI507/ΔF508 genotype to wt/ΔF508 in the 17-16-C1 iPSCs now resulted in restored expression of the 170 kDa CFTR protein. As expected this mature glycoform was also glycosylase sensitive.

Figure 3.26 Expression profile of differentiated cells after 19 days indicate differentiation towards early lung and thyroid endodermal lineages. RNA was isolated from day 19 differentiated cells derived from uncorrected iPS Clone 17, corrected and cre-excised Clone 17-16-C1, or WA09. Expression levels measured by qRT-PCR indicate upregulation of pan-endodermal marker *FOXA2*, Thyroid marker *PAX8*, and several lung markers (*CFTR, NKX2.1, SOX9, P63, and FOXP2*) as well as downregulation of pluripotency marker *OCT4*, indicating differentiation towards early anterior foregut lineages. As a positive control, RNA extracts from fetal lung, fetal thyroid, and Calu-3 cells were included where applicable. Note that y-axis is on a log scale

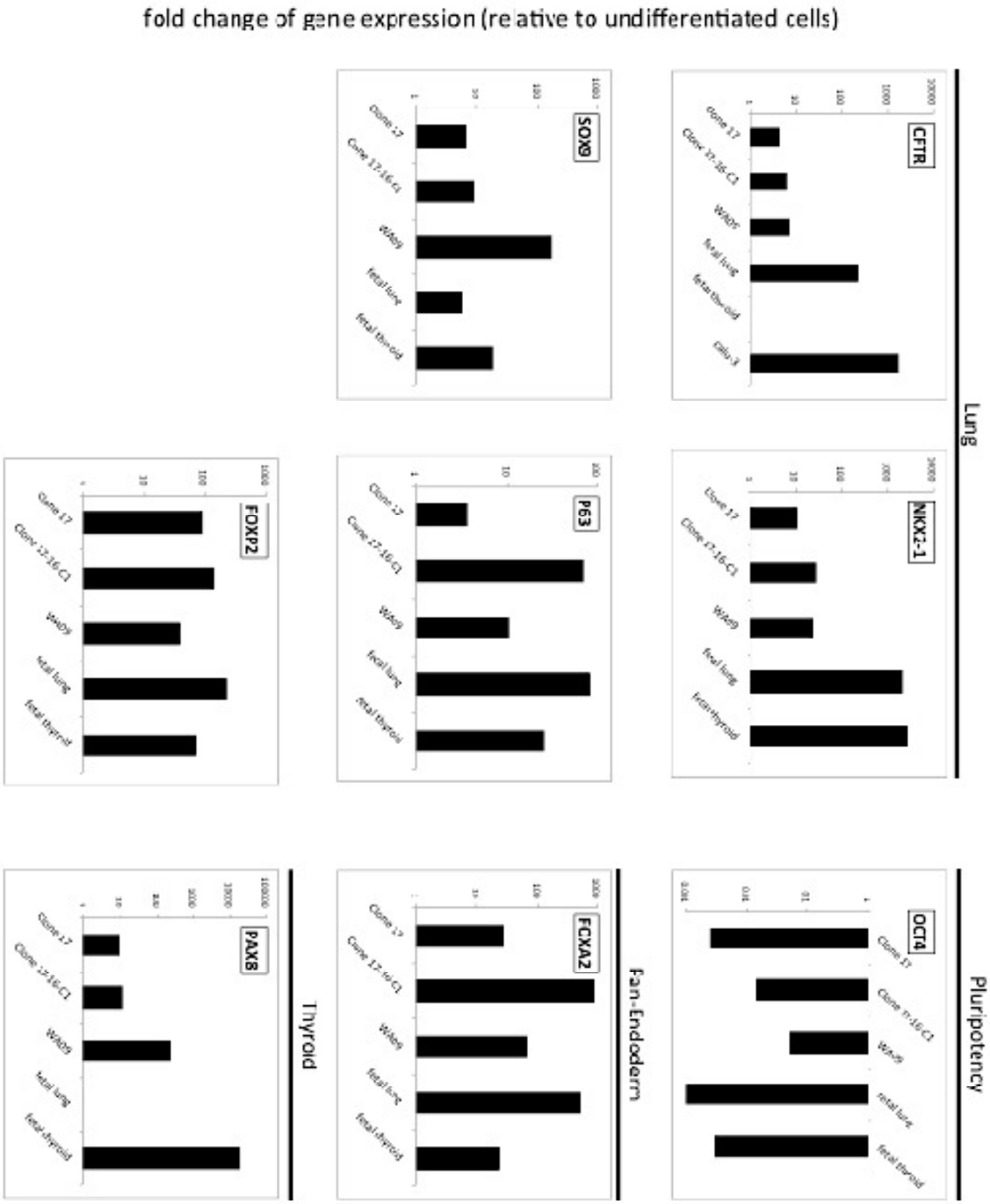

fold change of gene expression (relative to undifferentiated cells)
Lung
Pluripotency
Foregut-Endoderm
Thyroid
CFTR
NKX2-1
OCT4
SOX9
P63
FOXP2
FOXA2
PAX8
Clone 17
Clone 17-16-C1
WA09
fetal lung
fetal thyroid
Calu-3

FIGURE 3.27 Differentiated cells express CFTR RNA with expected sequence. Sequencing of *CFTR* mRNA isolated from CFTR-expressing Calu-3 cells and from differentiated cells derived from WA09 show expected wild-type sequence, whereas when isolated from differentiated uncorrected iPSCs (Clone 17) both ΔI507 and ΔF508 deletions are observed. Importantly, cells differentiated from corrected, Cre-excised iPS (Clone 17-16-C1) produce mRNA with one wt and one ΔF508 allele, corresponding to carrier state.

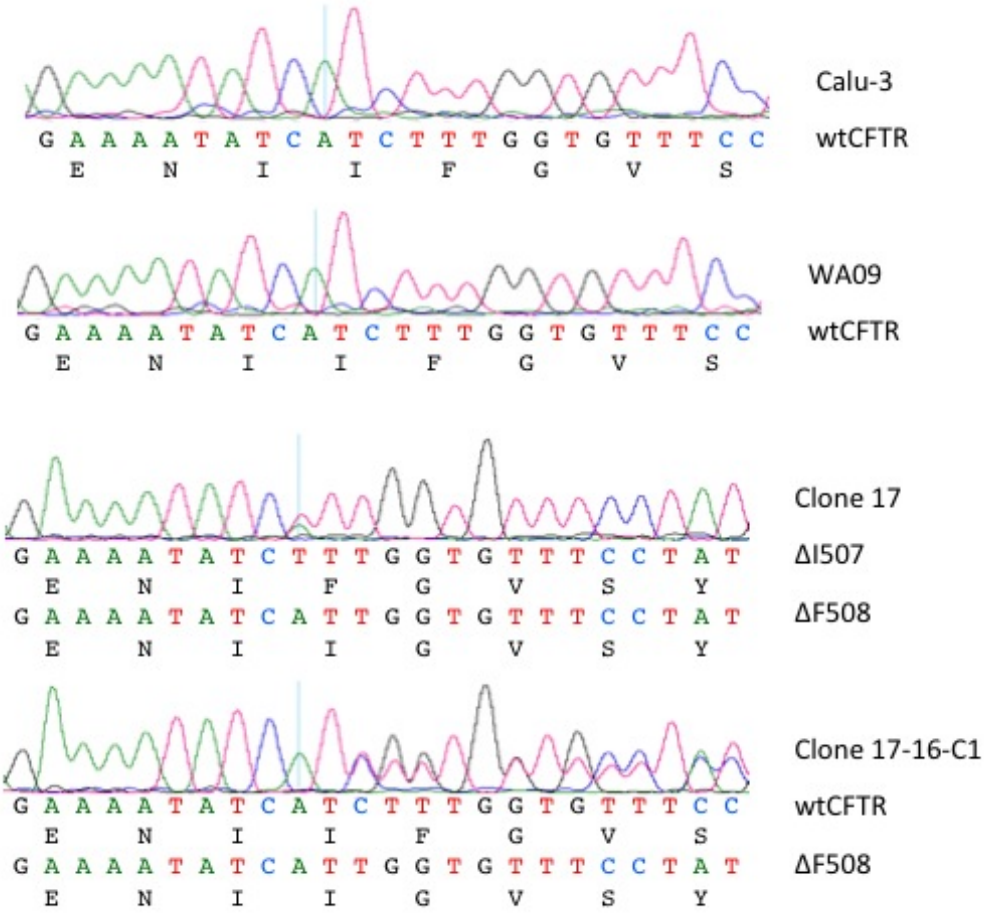
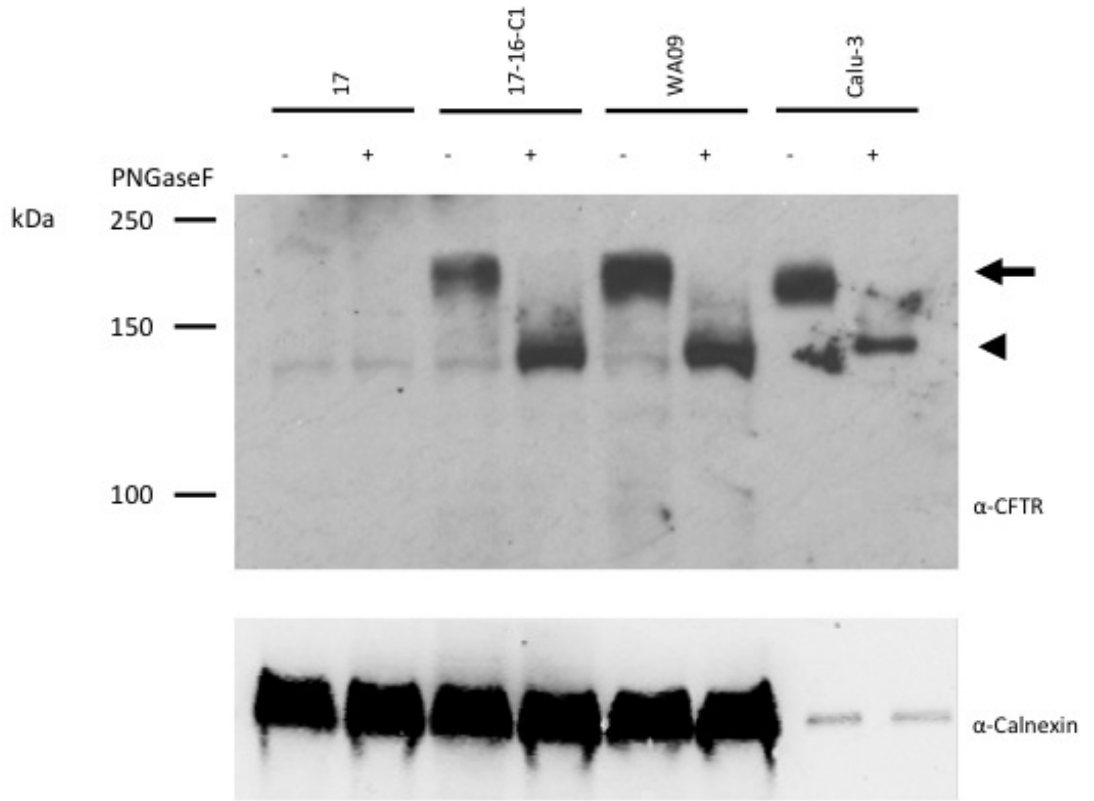


FIGURE 3.28 Differentiated corrected iPSCs express mature CFTR. As expected, immunoblotting of wild-type Calu-3 cells and those differentiated from WA09 show a predominance of the fully-glycosylated, mature form of CFTR of ~170 kDa (arrow). This mature isoform is sensitive to deglycosylation by PNGase F, an *N*-glycosidase, (+) resulting in a reduction of molecular weight to ~140 kDa corresponding to an immature CFTR (arrowhead). Uncorrected iPS (17), on the other hand, does not generate mature CFTR protein but show low level of the immature isoform, which is quickly degraded before reaching structural maturity. However, correction of one allele, as in cells derived from 17-16-C1, restores the fully-glycosylated, mature form that shows similar sensitivity to PNGaseF as did Calu-3 and cells derived from WA09. Calnexin was used as a loading control.



Chapter 4: Discussion

The ability to reprogram pluripotent stem cells directly from patient skin or other somatic cells and to edit the genome in a site-specific manner in these induced pluripotent stem cells, has opened up an avenue for treatment of monogenic diseases like Cystic Fibrosis. CF is an autosomal recessive disease for which current treatments provide only short-term alleviation of symptoms and attempts at gene therapy have failed (largely due to inefficacy of targeting appropriate cell types and issues with immune responses against repeated delivery of the virus). We thus sought to apply these two novel technologies to CF as groundwork for alternative cellular and gene therapy – the goal being to provide an autologous, long-term treatment that not only addresses symptoms but also is curative in essence. We have shown that zinc finger nuclease-mediated homologous directed repair can be used to correct two distinct *CFTR* mutations (ΔI507 and ΔF508) in iPSCs; for these studies we utilized either iPSCs derived by retroviral-mediated integration of the reprogramming factors or iPSCs generated in a transgene-free method by another laboratory. Furthermore, we have demonstrated that such correction was made with no apparent off-target effects in exon sequences as indicated by complete genome and whole exome sequencing, as well as bioinformatic analysis of potential off-target sites. Significantly, we prove that differentiated cells derived from the corrected iPSCs produce the fully glycosylated, mature form of CFTR.

4.1 Generation of CF-iPSCs

When we first began to generate iPSCs, we relied on viral delivery of the reprogramming factors originally used by Yamanaka's and Daley's groups (*45*, *46*). We initially attempted to use lentivirus to deliver the Thomson reprogramming factors (*134*),

but succeeded only with retrovirus expressing the 4 Yamanaka factors plus NANOG. As reported by others, we found that the key to reprogramming was having high titers of the reprogramming vectors. Given that reprogramming is an inefficient process, it is crucial to maximize the number of starting cells receiving all reprogramming factors. Furthermore, Yamanaka and colleagues showed that iPS generated using retrovirus had integrated three to six copies of each factor, suggesting that a certain combination and/or level of induction is required for full reprogramming (*45*). Thus, to achieve high titers of our retrovirus, we adapted Park and Daley's protocol for VSV-G pseudotyped retroviral vectors, which allowed us to concentrate the virus production by ultracentrifugation and thus transduce at a high multiplicity of infection (*120*).

Upon delivery of the factors and selection for reprogramming cells, isolation of those cells became crucial in creating clonal iPSC colonies. In their detailing of the reprogramming process, Lowry et al. described the emergence of two waves of colonies. The first wave was found to be unexpandable (*56*). Significantly, they showed that this first wave consisted of colonies transduced only with *OCT4* and *C-MYC* retroviruses. It was from the second wave of colonies, which they found to be TRA1-81 positive, that they isolated fully reprogrammed colonies. We also observed two waves of colonies arising from our experiments. In isolating our iPSCs, we initially relied on morphology alone, looking for the morphologies as described by Lowry et al. We observed many colonies that fit the human ESC colony morphology, but the first wave of colonies did not expand. Furthermore, we found that, when we used live cell staining for antigen TRA-1-60 as another criteria, morphology and TRA-1-60 positivity did not always coincide (FIGURE 3.2). Many of these colonies that did not initially resemble ESC

colonies but expressed TRA-1-60 began to take on a morphology indistinguishable from ESCs as they were isolated and further passaged. Thus, we found live-cell staining for TRA-1-60 to be a more reliable indicator of reprogramming.

4.2 ZFN-mediated correction of CF-iPSCs

Having generated CF-specific iPSCs that were ΔI507/ΔF508 compound heterozygous, our objective was to then generate a patient-specific iPSC line for CF that was corrected for the disease-conferring mutation with as little change to the underlying cellular genome as possible. While the method we used for reprogramming resulted in multiple integrations (thus compromising the integrity of the cell's genome) this can be circumvented for future clinical applications with Cre-excisable polycistronic or non-integrating reprogramming vectors, as previously discussed. When it came to correcting the CF mutation, we chose site-specific approach correction of the endogenous mutation rather than integration of the wt *CFTR* cDNA in a safe harbor or random integration site; our objective was that regulation of the corrected *CFTR* gene would still be under the control of endogenous *CFTR* gene regulatory elements. To facilitate site-specific gene targeting, we utilized ZFNs bioengineered to create a DSB in exon 10 of the CF locus, upstream of the ΔI507/ΔF508 mutations. The ZFNs were obtained through a collaboration with Sangamo Biosciences, who not only bioengineered the ZFNs but provided much insight and guidance on optimizing the delivery of the ZFNs as well as on construction of a donor suitable for our purpose.

There were several considerations in the design and construction of the donor. The first thing we considered was the length of the homology arms. While 5 to 10 kb of total homology is typically recommended for traditional HR in mouse or human (*135*),

zinc finger nuclease-mediated HR require significantly shorter homology sequences (on the order of 1-1.5 kb total). For example, Moehle et al. demonstrated that both 1 kb and 7.7 kb insertions, each flanked by 750 kb homology arms, were incorporated by ZFN-mediated HDR at an efficiency of about 5% (*106*). We, thus, decided to construct our donor similarly with approximately 1.6 kb of total homology. This would, based on the location of the DSB, encompass all of exon 10 with enough room on each end to include splice signal sequences and flanking intron sequences. Additionally, because the homology arms would also include the ZFN recognition sequence, we introduced two single base pair mutations that would alter the sequence recognized by the right ZFN without altering the encoded amino acid sequence.

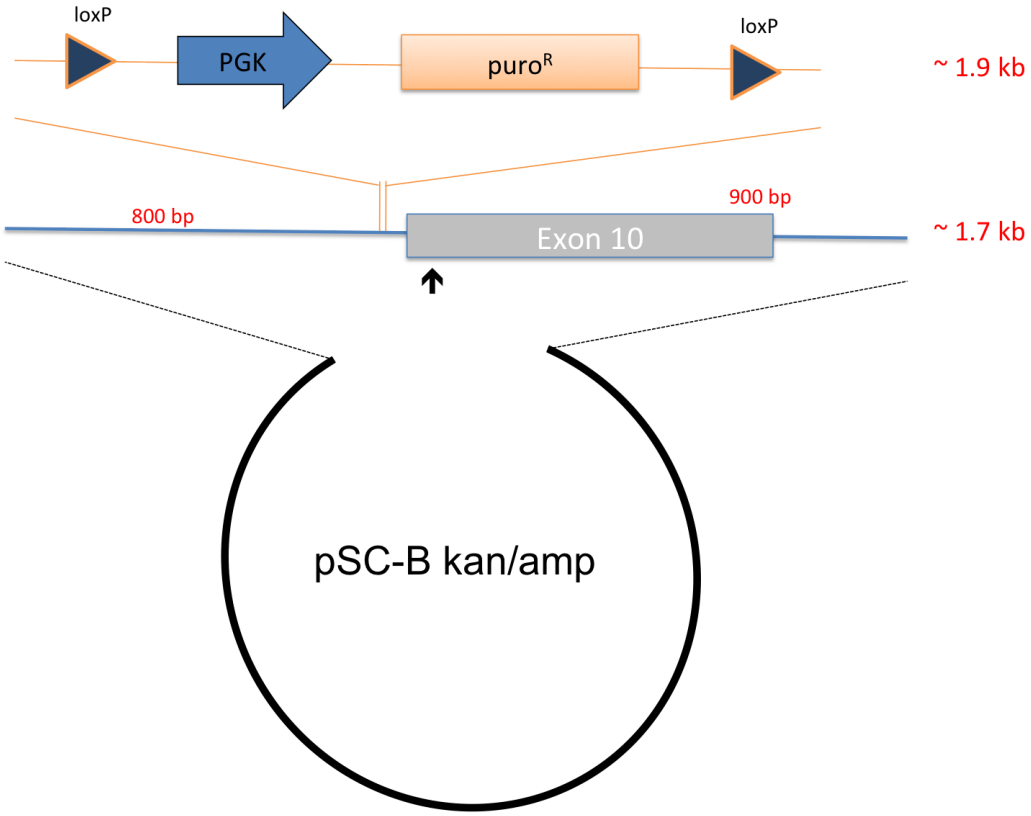
Though Soldner et al. have demonstrated the ability to integrate genes in iPSCs without the use of drug selection (*53*), we opted to take a drug-selection based strategy to afford us the ability to quickly assess and select for integration events. We found that puromycin, more so than neomycin, provided swift and efficient selection to minimize the chances of growth of cells without integration. By utilizing a positive/negative selectable marker conferring puromycin resistance and thymidine kinase (TK) sensitivity, we were able to distinguish integration events by positive selection of puromycin resistance and, later, Cre-excision events by FIAU resistance. In addition, we wanted our selection marker to be excisable and so opted for one that was flanked by loxP sites, similar to the excisable selection marker utilized in correction of sickle cell anemia (*114*). By Cre-mediated recombination, the selection marker could be excised leaving just one loxP site of 34 bases still integrated. For future adaptation of our methods, if a selection

marker is used, we suggest that they be flanked with *piggyBac* transposon sequences, which allows for excision without leaving any residual sequences.

In order to keep the homology arms of almost equal lengths on both sides of the ZFN target site, we initially put the selection marker in intron 9, approximately 100 bases upstream of exon 10 (FIGURE 4.1). In our first few experiments using this construct to correct the ΔI507/ΔF508 iPS, the donor was and ZFNs were delivered by electroporation using protocols not yet optimized. We were able to isolate a sufficient number of puromycin-resistant colonies, a small number of which showed evidence of targeted integration by PCR. However, analysis of the sequence for these putative targeted integrations revealed no correction of either the ΔI507 or ΔF508 alleles. This suggests that HR did not extend far enough to include the corrective sequences from the donor to repair the ΔI507 or ΔF508 mutations. Due to the low number of puromycin-resistant clones from samples nucleofected with donor alone, it is unlikely that these puromycin-resistant yet not corrected clones resulted from random integration events; however, we did not investigate these clones any further and thus cannot conclusively exclude this possibility.

Our collaborator, Dr. Michael Holmes from Sangamo Biosciences, then suggested placing the selection cassette in intron 10. While we were not able to preserve the equal size of homology arms, this would place any changes in the donor relative to the endogenous *CFTR* sequences on one side of the DSB and, importantly, require any HR event that integrated the selection cassette to also include the corrective sequences. It was with this donor, in which the selection cassette was placed in intron 10, that we were able to obtain colonies with targeted correction.

FIGURE 4.1 Donor with selectable cassette in intron 9.



Another consideration was whether to correct before or after reprogramming. Because 1) correction would require expansion both before and after Cre-excision and 2) reprogrammed cells can be passaged for a long time without senescence, we opted to reprogram the cells first and then correct them. Another benefit of this approach is that we could generate banks of both uncorrected and corrected iPSCs that were clones of one another except for the changes introduced during the correction. Had we corrected first, it would be highly unlikely that we would have uncorrected and corrected reprogrammed cells that had the same reprogramming profile (i.e. number and location of insertions of reprogramming factors). There are, however, cases in which correction before reprogramming would be beneficial. Such is the case with Fanconi anaemia, where cells show genomic instability and are predisposed to apoptosis (*52*). In this case, investigators found it beneficial to correct before or at the same time as reprogramming.

4.3 Evidence of allele-preferred targeting

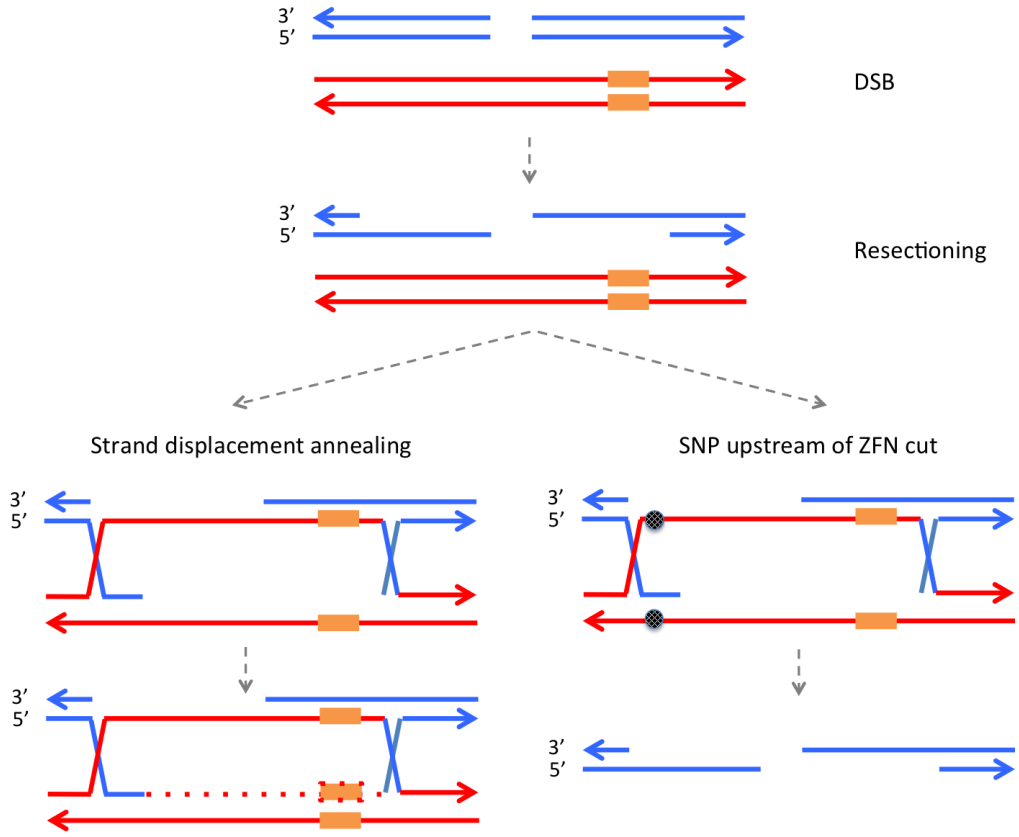
Sequencing of colonies that showed targeted integration of the donor with the selection cassette in intron 10 (Donor 1) revealed significant preferential targeting of one allele over the other (6:1 targeting of ΔI507 over ΔF508; FIGURE 3.20). Given that the number of changes in the donor relative to the two endogenous alleles was the same (i.e. insertion of 3 bp), we anticipated nearly equal targeting of both alleles. Demonstrated CFTR mRNA expression arising from both the ΔI507 and ΔF508 alleles in the targeted Clone 17 iPSCs suggested that this allele preferential targeting was not due to different transcriptional activity (e.g., if ΔI507 had been transcribed but ΔF508 silent). Further investigation into the sequences of each allele, specifically those sequences that were included in the donor, revealed that there was, in fact, a single base pair difference 76 bp

upstream of exon 10 (this same SNP was also found in the originating fibroblast cells and is a documented SNP). Where the endogenous ΔI507 allele had an adenosine, the ΔF508 allele had a guanine; the original donor we had constructed had an adenosine, thus matching the ΔI507 allele at this position. Changing the donor such that it matched the ΔF508 allele at this position resulted in a reversal of preference with 1:3.7 (ΔI507:ΔF508) showing targeting (Donor 2, FIGURE 3.20. Though surprising that a single base pair difference could have such drastic effects, our finding corroborates past experiments by Waldman and Liskay in 1988 (*109*) and later by Elliot et al. (*110*); both groups found that critical to efficiency of HR is the amount of uninterrupted homology. Collectively, these findings would suggest, then, that a break in contiguous homology introduced by a SNP 76 bases upstream of exon 10 (87 bp from the spacer region) would not provide a sufficient stretch of uninterrupted homology for efficient repair of a DSB (FIGURE 4.2). These findings underscore the need to closely inspect and match the sequences in the donor to those present in the targeted cell line(s), subtle changes could account for low targeting efficiencies. Additionally, this ability to preferentially target a specific allele might be advantageous in cases where targeting of one allele vs the other is essential in restoring the gene, as in dominant diseases like Huntington's disease or myotonic distrophy.

4.4 Targeting and correction of one vs two alleles

In attempting to correct either the compound heterozygous CF-iPSCs or the homozygous ΔF508 transgene-free iPSCs, we achieved targeting and correction of just one allele. There are several factors that we believe would be necessary to achieve biallelic correction (though, again, not necessary for a recessive disease like Cystic

FIGURE 4.2 Mechanism for preferential targeting. In homology directed repair, there is a resection of the 5' ends of each cut strand. Then, in strand displacement annealing the cut strand invades the template and is repaired in 5' to 3' direction using the antisence donor strand as a template. If a SNP exists between the donor and the template just upstream of the DSB in the genome but beyond the portions resectioned, this SNP can inhibit proper invasion and, thus, strand displacement annealing.



Fibrosis). First we would have to increase the efficiency of cleaving both alleles by the ZFNs, since this is a requisite event for targeting. Sequencing of the unmodified alleles in our CF-iPS cells targeted with Donor 1 (FIGURE 3.20) revealed that that only 1 out of the 21 clones sequenced showed evidence of cutting as evidenced by introduced mutations at the cut site by NHEJ (data not shown). This suggests that the ZFNs used in our experiments rarely cut both alleles. Second, to achieve biallelic targeting we would also have to select a donor(s) that matched each of the targeted alleles. In the case of the homozygous ΔF508 cells (sequencing of which showed that both alleles had a G at the position in intron 9), Donor 2 should be used for optimal targeting. For targeting of our compound heterozygous CF-iPSCs, the use of both donors (the Donor 1 for ΔI507 and the Donor 2 for ΔF508) would increase the possibility of sucessful targeting of both alleles. Finally, to select for biallelic targeting, we would suggest the use of a double selection method, one specific for each allele. For example, one of the donors would confer puromycin resistance while the other confers neomycin resistance. Selection of puromycin and neomycin resistant colonies would then enrich for those with targeted integration of both donors.

4.5 Demonstration of restored CFTR protein expression by targeted differentiation of corrected iPSCs

Since CF-related complications in the lung account for 90% of CF morbidity and mortality, we explored differentiation of our iPSCs into lung cells with the goal of demonstrating correction-mediated restoration of expression of the CFTR protein. Fortunately, there has been much headway in elucidating the pathways involved in lung organogenesis and how to translate this into a differentiation protocol for human ESCs.

Building upon these studies, we have been able to differentiate several pluripotent stem cells and demonstrate CFTR expression.

Though there is still some question of the specific lung cell type(s) that express CFTR, we initially assumed we would have to generate fully differentiated, mature cells (like ciliated cells) to observe CFTR protein. When we finally began differentiation experiments, there were reports of differentiation of pluripotent stem cells into definitive endoderm, the precursor to the anterior (lung and thyroid) and posterior (intestine, liver, pancreas) foregut, in a way that attempts to recapitulate normal embryonic development: by inducing Nodal/Activin signaling, Activin A treatment can effectively generate definitive endoderm from mouse and human ESCs (*125*, *131*). There were also two reports of the generation of respiratory epithelia-like tissue from mouse and human ESCs (*13*, *132*). These latter reports relied on spontaneous differentiation guided by culture conditions that mimic the air-liquid interface (ALI) of the lung. When we tried to reproduce both the protocols for definitive endoderm and for lung epithelial tissue, we did observe upregulation of CFTR gene expression by qPCR (FIGURE 3.24), sequencing of which revealed appropriate transcriptional regulation of both the corrected and unmodified alleles. We also observed, in the Activin A experiments, expression of Sox17, a gene expressed in definitive endoderm. Unfortunately, though we were able to detect *CFTR* mRNA, we were unable to detect CFTR protein.

Because these attempts by Coroux et al. and Van Haute et al. rely on spontaneous differentiation, it is questionable whether the tissue generated is purely lung tissue. Furthermore, it is still unclear whether these reports are reproducible. We, thus, put our efforts into more step-wise differentiation approaches like that employed in the derivation

of definitive endoderm. The belief and hope is that by taking an approach that attempts to mimic the signaling pathways that direct lung organogenesis, we can then begin to establish protocols that will efficiently and reproducibly generate the specific cell types in which we are interested, including both CFTR-expressing cells as well as progenitors to such cells. Two groups have paved the way for such step-wise differentiation of definitive endoderm into anterior foregut and subsequently early lung progenitors (*79*, *80*). Taking their cues from normal development, both groups emphasized the need to generate, from definitive endoderm, the primordial progenitors to the anterior foregut. This stage, which is marked by NKX2.1 expression, requires anteriorization due to inhibition of BMP and TGFβ signaling and then ventralization. Longmire et al. went further and showed that murine ESC-derived Nkx2.1^{+} cells were capable of further differentiation into more mature lung and thyroid lineages as determined by detection of lung- and thyroid-specific mRNAs. In collaboration with the Kotton group, we assessed whether their step-wise differentiation of murine ESC-derived Nkx2.1-expressing cells could be translated for differentiation of human ESCs and iPSCs, with the hopes of generating CFTR-expressing cells, as well as other lung and thyroid lineages.

We began to observe significant upregulation of *CFTR* mRNA (relative to positive controls) upon anteriorization and ventralization of definitive endoderm (FIGURE 3.27). We, thus, explored whether we could detect even low levels of CFTR protein at this point, and we were able to detect CFTR protein after 19 days of differentiation (FIGURE 3.28). Notably, when derived from uncorrected iPSCs (17), only low levels of the immature, core-glycosylated form of CFTR is observed; however, when derived from corrected, Cre-excised iPSCs (17-16-C1), both the core- and fully-

glycosylated forms are observed. Sensitivity to deglycosylation by *N*-glycosidase F of the protein detected from the 17-16-C1 derived cells, further confirms that the upper band is the fully glycosylated, mature form of CFTR. Patterns observed from the 17-16-C1 derived cells are indistinguishable from that observed from similarly differentiated WA09 cells and from CFTR-expressing Calu-3 cells. It must be noted that mRNA expression profile of the differentiated cells suggest that we have not generated just lung and thyroid progenitor(s) but rather a heterogenous population. We thus cannot make any conclusion about whether these are lung CFTR-expressing cells. Nonetheless, these results still conclusively indicate that these CF-iPSCs that were corrected by ZFN-mediate HDR can differentiate into cells that produce the mature, fully glycosylated CFTR.

Recently, two groups have used ZFNs to restore or induce *CFTR* gene expression. Lee et al. used ZFNs to perform site-specific correction of the ΔF508 mutation in bronchial epithelial cells (*136*). However, they failed to demonstrate restored CFTR protein expression. Furthermore, they did not clonally isolate the corrected cells. Ramalingam et al. reported integration of the CFTR transcription unit into the *CCR5* safe harbor locus in wild-type iPSCs with constitutive CFTR expression already in undifferentiated iPSCs (*137*). We report here the first documentation of site-specific correction of the *CFTR* gene in CF-specific iPSCs, as demonstrated at both the genetic and protein level. This is, additionally, the first demonstration of allele-preferred targeting in human cells, an observation that may help improve targeting efficiencies and assist in correcting dominant mutations.

4.6 Future direction

As already noted above, there are areas in which improvements need to be made before our demonstration of site-specific correction of CF-iPSCs is clinically applicable (eg. the effects of the reprogramming process and prolonged culture of iPSCs on its genomic and epigenomic stability; improved efficacy of site-specific, DSB-induced homologous recombination; and derivation of pure and functional populations of lung epithelial cells from pluripotent stem cells). Discussion of future work and direction, then, will focus primarily on 1) assessing appropriate cellular localization of CFTR and restored protein function and 2) how these cells can be used for cellular therapy of CF.

We have shown that the genetic correction made to the CF-specific iPSCs is appropriately translated as we are able to detect CFTR protein by western blot. As the absence of CFTR in ΔI507 and ΔF508 patients is due to misfolding of the protein, arrest of protein maturation, and subsequent degradation of the immature protein, it follows that genetic correction of the mutation that results in expression of mature CFTR protein would result in proper trafficking of the protein. Immunostaining of CFTR at the apical surface of CFTR-expressing cells is sufficient to demonstrate appropriate localization of the restored CFTR protein. To unequivocally prove apical expression of the protein, we would need to generate a polarized culture.

If the corrected CFTR is appropriately trafficked to the apical membrane, we should be able to detect restored chloride transport function by electrophysiological assays, such as patch-clamp or Using chamber. As both techniques will require quite a bit of expertise that we do not have, we have been actively considering another option: chloride flux assays. Using fluorescent probes sensitive to iodide, which is also CFTR permeable, intracellular fluorescence is used as an indicator of CFTR activity upon

activation with forskolin and inhibition upon treatment with CFTR-specific inhibitor $CFTR_{inh}$-172 (*138-140*). Unlike Patch-clamp, this assay seems to be more straightforward. Furthermore, unlike Ussing chamber, the assay can be performed on dissociated cells, thus polarity is not a requirement. If sensitivity is an issue, we might consider radioactive tracers, which are more sensitive to small fluxes.

With proof that functional CFTR-expressing cells can be generated from the iPSCs, we are one step closer to making cellular therapy a viable option for treatment of CF. Transplantation of CFTR-expressing cells derived from the corrected iPSCs would provide immediate treatment for CF. However, because these are terminally differentiated cells and the airway epithelium has a quick turnover rate (*1*), co-delivery of stem progenitors to those CFTR-expressing cells could treat provide long-term treatment of the disease. Unfortunately, as aforementioned, neither the CFTR-expressing cells nor the progenitor(s) to those cells have been unequivocally identified. Considerations for means of delivering the cells will also not be trivial. For example, the CF lung will already exhibit damage and viscous mucous, which may inhibit engraftment of transplanted cells (as it inhibits efficient delivery of viral gene therapy agents).

As an alternative, many labs are exploring the possibility of using a lung scaffold to help direct differentiation of pluripotent and multipotent cells (*141-143*). For example, Niklason and collegues decellularized a rat lung resulting in an extracellular matrix scaffold that maintained the branching structure of the lung (*144*). By seeding the decellularized lung with pulmonary epithelium and vascular endothelium, they demonstrated the cells were capable of repopulating the respiratory compartments and, notably, the engineered lung was capable of gas exchange when transplanted into rats *in*

vivo. These and similar works have sparked interest aimed at determining whether these decellularized lung scaffolds that can be populated by the corrected iPSCs or, more likely, early lung progenitors ($NKX2.1^{+}$) cells derived from corrected iPSCs to generate a functional lung for transplantation.

As quickly as advances in reprogramming (to make iPSCs more clinically safe) are being made, the limiting step for clinical application of these cells, I believe, lies with directed differentiation of target cells (ie. the affected cell types or progenitors to those cells). Nonetheless, even in the time it has taken to write this dissertation, we have learned that there has been progress made in further differentiating iPSCs into lung cells and the use of these cells in recellularization of lung scaffolds. Furthermore the Japanese Ministry of Health has just approved a clinical study into treatment of age-related macular degeneration using iPSC-derived retinal pigmented epithelium (145) and the US FDA has approved a phase I clinical trial on the safety and tolerability of iPS-derived retinal pigmented epithelium in patients with Stargardt's Macular Dystrophy (clinical trial identifier NCT01345006). Though there are concerns about the readiness of iPSCs and iPSC-derived cells for human use, these pilot studies is just one indication of the speed at which research, and even clinical translation, in this promising area of regenerative medicine is progressing.

References

1. Mueller, C., and Flotte, T. R. (2008) Gene therapy for cystic fibrosis, *Clinic Rev Allerg Immunol 35*, 164–178.
2. Ratjen, F., and Döring, G. (2003) Cystic fibrosis, *Lancet 361*, 681–689.
3. Riordan, J. R. (1999) Cystic fibrosis as a disease of misprocessing of the cystic fibrosis transmembrane conductance regulator glycoprotein, *Am. J. Hum. Genet. 64*, 1499–1504.
4. Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., Buchwald, M., and Tsui, L. C. (1989) Identification of the cystic fibrosis gene: genetic analysis, *Science 245*, 1073–1080.
5. Champigny, G., Imler, J. L., Puchelle, E., Dalemans, W., Gribkoff, V., Hinnrasky, J., Dott, K., Barbry, P., Pavirani, A., and Lazdunski, M. (1995) A change in gating mode leading to increased intrinsic Cl- channel activity compensates for defective processing in a cystic fibrosis mutant corresponding to a mild form of the disease, *The EMBO Journal 14*, 2417–2423.
6. Orozco, L., Friedman, K., Chávez, M., Lezana, J. L., Villarreal, M. T., and Carnevale, A. (1994) Identification of the I507 deletion by site-directed mutagenesis, *Am. J. Med. Genet. 51*, 137–139.
7. Super, M., and Schwarz, M. J. (1992) Mutations of the cystic fibrosis gene locus within the population of the Northwest of England, *Eur. J. Pediatr. 151*, 108–111.
8. Qu, B. H., Strickland, E. H., and Thomas, P. J. (1997) Localization and suppression of a kinetic defect in cystic fibrosis transmembrane conductance regulator folding, *Journal of Biological Chemistry 272*, 15739–15744.

9. Wang, F., Zeltwanger, S., Hu, S., and Hwang, T. C. (2000) Deletion of phenylalanine 508 causes attenuated phosphorylation-dependent activation of CFTR chloride channels, *J. Physiol. (Lond.) 524 Pt 3*, 637–648.

10. Farmen, S. L., Karp, P. H., Ng, P., Palmer, D. J., Koehler, D. R., Hu, J., Beaudet, A. L., Zabner, J., and Welsh, M. J. (2005) Gene transfer of CFTR to airway epithelia: low levels of expression are sufficient to correct Cl- transport and overexpression can generate basolateral CFTR, *Am. J. Physiol. Lung Cell Mol. Physiol. 289*, L1123–30.

11. Qu, B. H., and Thomas, P. J. (1996) Alteration of the cystic fibrosis transmembrane conductance regulator folding pathway, *Journal of Biological Chemistry 271*, 7261–7264.

12. Vankeerberghen, A., Cuppens, H., and Cassiman, J.-J. (2002) The cystic fibrosis transmembrane conductance regulator: an intriguing protein with pleiotropic functions, *J. Cyst. Fibros. 1*, 13–29.

13. Puchelle, E., Gaillard, D., Ploton, D., Hinnrasky, J., Fuchey, C., Boutterin, M. C., Jacquot, J., Dreyer, D., Pavirani, A., and Dalemans, W. (1992) Differential localization of the cystic fibrosis transmembrane conductance regulator in normal and cystic fibrosis airway epithelium, *Am. J. Respir. Cell Mol. Biol. 7*, 485–491.

14. Koch, C., and Høiby, N. (1993) Pathogenesis of cystic fibrosis, *Lancet 341*, 1065–1069.

15. Cohen, T. S., and Prince, A. (2012) Cystic fibrosis: a mucosal immunodeficiency syndrome, *Nat Med 18*, 509–519.

16. Lubamba, B., Dhooghe, B., Noel, S., and Leal, T. (2012) Cystic fibrosis: insight into CFTR pathophysiology and pharmacotherapy, *Clinical Biochemistry 45*, 1132–1144.

17. Anson, D. S., Smith, G. J., and Parsons, D. W. (2006) Gene therapy for cystic fibrosis airway disease- is clinical success imminent?, *Curr Gene Ther 6*, 161–179.

18. Cystic Fibrosis Foundation. Cystic Fibrosis Foundation Patient Registry. 2011 annual data report. Bethesda, Maryland. Available from: http:// www.cff.org/treatments/CareCenterNetwork/PatientRegistryReport/

19. Quon, B. S., and Aitken, M. L. (2012) Cystic fibrosis: what to expect now in the early adult years, *Paediatric Respiratory Reviews 13*, 206–214.

20. Cheng, S. H., Gregory, R. J., Marshall, J., Paul, S., Souza, D. W., White, G. A., O'Riordan, C. R., and Smith, A. E. (1990) Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis, *Cell 63*, 827–834.

21. Sato, S., Ward, C. L., Krouse, M. E., Wine, J. J., and Kopito, R. R. (1996) Glycerol reverses the misfolding phenotype of the most common cystic fibrosis mutation, *Journal of Biological Chemistry 271*, 635–638.

22. Dormer, R. L., Dérand, R., McNeilly, C. M., Mettey, Y., Bulteau-Pignoux, L., Métayé, T., Vierfond, J. M., Gray, M. A., Galietta, L. J., Morris, M. R., Pereira, M. M., Doull, I. J., Becq, F., and McPherson, M. A. (2001) Correction of delF508-CFTR activity with benzo(c)quinolizinium compounds through facilitation of its processing in cystic fibrosis airway cells, *J. Cell. Sci. 114*,

4073–4081.

23. Kälin, N., Claass, A., Sommer, M., Puchelle, E., and Tümmler, B. (1999) DeltaF508 CFTR protein expression in tissues from patients with cystic fibrosis, *J. Clin. Invest. 103*, 1379–1389.

24. Kobayashi, M., Iida, A., Ueda, Y., and Hasegawa, M. (2003) Pseudotyped lentivirus vectors derived from simian immunodeficiency virus SIVagm with envelope glycoproteins from paramyxovirus, *J. Virol. 77*, 2607–2614.

25. Johnson, L. G., Olsen, J. C., Naldini, L., and Boucher, R. C. (2000) Pseudotyped human lentiviral vector-mediated gene transfer to airway epithelia in vivo, *Gene Ther. 7*, 568–574.

26. Crystal, R. G., McElvaney, N. G., Rosenfeld, M. A., Chu, C. S., Mastrangeli, A., Hay, J. G., Brody, S. L., Jaffe, H. A., Eissa, N. T., and Danel, C. (1994) Administration of an adenovirus containing the human CFTR cDNA to the respiratory tract of individuals with cystic fibrosis, *Nat Genet 8*, 42–51.

27. van Heeckeren, A., Ferkol, T., and Tosi, M. (1998) Effects of bronchopulmonary inflammation induced by pseudomonas aeruginosa on adenovirus-mediated gene transfer to airway epithelial cells in mice, *Gene Ther. 5*, 345–351.

28. Johnson, L. G., Vanhook, M. K., Coyne, C. B., Haykal-Coates, N., and Gavett, S. H. (2003) Safety and efficiency of modulating paracellular permeability to enhance airway epithelial gene transfer in vivo, *Human Gene Therapy 14*, 729–747.

29. Flotte, T., Carter, B., Conrad, C., Guggino, W., Reynolds, T., Rosenstein, B., Taylor, G., Walden, S., and Wetzel, R. (1996) A phase I study of an adeno-

associated virus-CFTR gene vector in adult CF patients with mild lung disease, *Human Gene Therapy 7*, 1145–1159.

30. Wagner, J. A., Nepomuceno, I. B., Messner, A. H., Moran, M. L., Batson, E. P., Dimiceli, S., Brown, B. W., Desch, J. K., Norbash, A. M., Conrad, C. K., Guggino, W. B., Flotte, T. R., Wine, J. J., Carter, B. J., Reynolds, T. C., Moss, R. B., and Gardner, P. (2002) A phase II, double-blind, randomized, placebo-controlled clinical trial of tgAAVCF using maxillary sinus delivery in patients with cystic fibrosis with antrostomies, *Human Gene Therapy 13*, 1349–1359.

31. Stotland, P. K., Radzioch, D., and Stevenson, M. M. (2000) Mouse models of chronic lung infection with Pseudomonas aeruginosa: models for the study of cystic fibrosis, *Pediatr. Pulmonol. 30*, 413–424.

32. Greber, U. F., Willetts, M., Webster, P., and Helenius, A. (1993) Stepwise dismantling of adenovirus 2 during entry into cells, *Nat Biotechnol 75*, 477–486.

33. Summerford, C., and Samulski, R. J. (1998) Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions, *J. Virol. 72*, 1438–1445.

34. Duan, D., Yue, Y., Yan, Z., McCray, P. B., and Engelhardt, J. F. (1998) Polarity influences the efficiency of recombinant adenoassociated virus infection in differentiated airway epithelia, *Human Gene Therapy 9*, 2761–2776.

35. Moss, R. B., Rodman, D., Spencer, L. T., Aitken, M. L., Zeitlin, P. L., Waltz, D., Milla, C., Brody, A. S., Clancy, J. P., Ramsey, B., Hamblett, N., and Heald, A. E. (2004) Repeated adeno-associated virus serotype 2 aerosol-mediated cystic fibrosis transmembrane regulator gene transfer to the lungs of patients with

cystic fibrosis: a multicenter, double-blind, placebo-controlled trial, *Chest 125*, 509–521.

36. Johnson, L. G., Olsen, J. C., Sarkadi, B., Moore, K. L., Swanstrom, R., and Boucher, R. C. (1992) Efficiency of gene transfer for restoration of normal airway epithelial function in cystic fibrosis, *Nat Biotechnol*, 1992/09/01 *2*, 21–25.

37. Johnson, L. G., Boyles, S. E., Wilson, J., and Boucher, R. C. (1995) Normalization of raised sodium absorption and raised calcium-mediated chloride secretion by adenovirus-mediated expression of cystic fibrosis transmembrane conductance regulator in primary human cystic fibrosis airway epithelial cells., *J. Clin. Invest. 95*, 1377–1382.

38. Zhang, L., Button, B., Gabriel, S. E., Burkett, S., Yan, Y., Skiadopoulos, M. H., Dang, Y. L., Vogel, L. N., McKay, T., Mengos, A., Boucher, R. C., Collins, P. L., and Pickles, R. J. (2009) CFTR delivery to 25% of surface epithelial cells restores normal rates of mucus transport to human cystic fibrosis airway epithelium, *PLoS Biol 7*, e1000155.

39. Engelhardt, J. F., Yankaskas, J. R., Ernst, S. A., Yang, Y., Marino, C. R., Boucher, R. C., Cohn, J. A., and Wilson, J. M. (1992) Submucosal glands are the predominant site of CFTR expression in the human bronchus, *Nat Genet 2*, 240–248.

40. Engelhardt, J. F., Zepeda, M., Cohn, J. A., Yankaskas, J. R., and Wilson, J. M. (1994) Expression of the cystic fibrosis gene in adult human lung, *J. Clin. Invest. 93*, 737–749.

41. Doucet, L., Mendes, F., Montier, T., Delépine, P., Penque, D., Férec, C., and

Amaral, M. D. (2003) Applicability of different antibodies for the immunohistochemical localization of CFTR in respiratory and intestinal tissues of human and murine origin, *J. Histochem. Cytochem. 51*, 1191–1199.

42. Kreda, S. M., Mall, M., Mengos, A., Rochelle, L., Yankaskas, J., Riordan, J. R., and Boucher, R. C. (2005) Characterization of wild-type and deltaF508 cystic fibrosis transmembrane regulator in human respiratory epithelia, *Molecular Biology of the Cell 16*, 2154–2167.

43. Mall, M., Kreda, S. M., Mengos, A., Jensen, T. J., Hirtz, S., Seydewitz, H. H., Yankaskas, J., Kunzelmann, K., Riordan, J. R., and Boucher, R. C. (2004) The DeltaF508 mutation results in loss of CFTR function and mature protein in native human colon, *Gastroenterology 126*, 32–41.

44. Takahashi, K., and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, *Cell 126*, 663–676.

45. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors, *Cell 131*, 861–872.

46. Park, I.-H., Zhao, R., West, J. A., Yabuuchi, A., Huo, H., Ince, T. A., Lerou, P. H., Lensch, M. W., and Daley, G. Q. (2008) Reprogramming of human somatic cells to pluripotency with defined factors, *Nature 451*, 141–146.

47. Aasen, T., Raya, Á., Barrero, M. J., Garreta, E., Consiglio, A., González, F., Vassena, R., Bilic, J., Pekarik, V., Tiscornia, G., Edel, M., Boué, S., and Izpisúa Belmonte, J. C. (2008) Efficient and rapid generation of induced pluripotent stem

cells from human keratinocytes, *Nat Biotechnol 26*, 1276–1284.

48. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G., and Hochedlinger, K. (2008) Induced pluripotent stem cells generated without viral integration, *Science 322*, 945–949.

49. Hanna, J., Markoulaki, S., Schorderet, P., Carey, B. W., Beard, C., Wernig, M., Creyghton, M. P., Steine, E. J., Cassady, J. P., Foreman, R., Lengner, C. J., Dausman, J. A., and Jaenisch, R. (2008) Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency, *Cell 133*, 250–264.

50. Park, I.-H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M. W., Cowan, C., Hochedlinger, K., and Daley, G. Q. (2008) Disease-specific induced pluripotent stem cells, *Cell 134*, 877–886.

51. Mali, P., Ye, Z., Hommond, H. H., Yu, X., Lin, J., Chen, G., Zou, J., and Cheng, L. (2008) Improved efficiency and pace of generating induced pluripotent stem cells from human adult and fetal fibroblasts, *Stem Cells 26*, 1998–2005.

52. Raya, Á., Rodríguez-Pizà, I., Guenechea, G., Vassena, R., Navarro, S., Barrero, M. J., Consiglio, A., Castellà, M., Río, P., Sleep, E., González, F., Tiscornia, G., Garreta, E., Aasen, T., Veiga, A., Verma, I. M., Surrallés, J., Bueren, J., and Izpisúa Belmonte, J. C. (2009) Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells, *Nature 460*, 53–59.

53. Soldner, F., Laganière, J., Cheng, A. W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L. I., Myers, R. H., Lindquist, S., Zhang, L., Guschin, D., Fong, L. K., Vu, B. J., Meng, X., Urnov, F. D., Rebar, E. J., Gregory, P. D., Zhang, H. S., and Jaenisch, R. (2011) Generation of isogenic pluripotent stem

cells differing exclusively at two early onset Parkinson point mutations, *Cell 146*, 318–331.

54. Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., Plath, K., and Hochedlinger, K. (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution, *Cell Stem Cell 1*, 55–70.

55. Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H., and Jaenisch, R. (2008) Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells, *Cell Stem Cell 2*, 151–159.

56. Lowry, W. E., Richter, L., Yachechko, R., Pyle, A. D., Tchieu, J., Sridharan, R., Clark, A. T., and Plath, K. (2008) Generation of human induced pluripotent stem cells from dermal fibroblasts, *Proceedings of the National Academy of Sciences 105*, 2883–2888.

57. Okita, K., Ichisaka, T., and Yamanaka, S. (2007) Generation of germline-competent induced pluripotent stem cells, *Nature 448*, 313–317.

58. Wernig, M., Meissner, A., Cassady, J. P., and Jaenisch, R. (2008) c-Myc is dispensable for direct reprogramming of mouse fibroblasts, *Cell Stem Cell 2*, 10–12.

59. Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A. E., and Melton, D. A. (2008) Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds, *Nat Biotechnol 26*, 795–797.

60. Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B. E., and Jaenisch, R. (2007) In vitro reprogramming of

fibroblasts into a pluripotent ES-cell-like state, *Nature 448*, 318–324.

61. Chang, C.-W., Lai, Y.-S., Pawlik, K. M., Liu, K., Sun, C.-W., Li, C., Schoeb, T. R., and Townes, T. M. (2009) Polycistronic lentiviral vector for "hit and run" reprogramming of adult skin fibroblasts to induced pluripotent stem cells, *Stem Cells 27*, 1042–1049.

62. Shao, L., Feng, W., Sun, Y., Bai, H., Liu, J., Currie, C., Kim, J., Gama, R., Wang, Z., Qian, Z., Liaw, L., and Wu, W.-S. (2009) Generation of iPS cells using defined factors linked via the self-cleaving 2A sequences in a single open reading frame, *Cell Res 19*, 296–306.

63. Somers, A., Jean, J.-C., Sommer, C. A., Omari, A., Ford, C. C., Mills, J. A., Ying, L., Sommer, A. G., Jean, J. M., Smith, B. W., Lafyatis, R., Demierre, M.-F., Weiss, D. J., French, D. L., Gadue, P., Murphy, G. J., Mostoslavsky, G., and Kotton, D. N. (2010) Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette, *Stem Cells 28*, 1728–1740.

64. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009) Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome, *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci. 85*, 348–362.

65. Yusa, K., Rad, R., Takeda, J., and Bradley, A. (2009) Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon, *Nat Meth 6*, 363–369.

66. Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I. I., and Thomson,

J. A. (2009) Human induced pluripotent stem cells free of vector and transgene sequences, *Science 324*, 797–801.

67. Zhou, H., Wu, S., Joo, J. Y., Zhu, S., Han, D. W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y., Siuzdak, G., Schöler, H. R., Duan, L., and Ding, S. (2009) Generation of induced pluripotent stem cells using recombinant proteins, *Cell Stem Cell 4*, 381–384.

68. Laurent, L. C., Ulitsky, I., Slavin, I., Tran, H., Schork, A., Morey, R., Lynch, C., Harness, J. V., Lee, S., Barrero, M. J., Ku, S., Martynova, M., Semechkin, R., Galat, V., Gottesfeld, J., Izpisúa Belmonte, J. C., Murry, C., Keirstead, H. S., Park, H.-S., Schmidt, U., Laslett, A. L., Muller, F.-J., Nievergelt, C. M., Shamir, R., and Loring, J. F. (2011) Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture, *Cell Stem Cell 8*, 106–118.

69. Hussein, S. M., Batada, N. N., Vuoristo, S., Ching, R. W., Autio, R., Närvä, E., Ng, S., Sourour, M., Hämäläinen, R., Olsson, C., Lundin, K., Mikkola, M., Trokovic, R., Peitz, M., Brüstle, O., Bazett-Jones, D. P., Alitalo, K., Lahesmaa, R., Nagy, A., and Otonkoski, T. (2011) Copy number variation and selection during reprogramming to pluripotency, *Nature 471*, 58–62.

70. Martins-Taylor, K., and Xu, R.-H. (2012) Concise review: Genomic stability of human induced pluripotent stem cells, *Stem Cells 30*, 22–27.

71. Baker, D. E. C., Harrison, N. J., Maltby, E., Smith, K., Moore, H. D., Shaw, P. J., Heath, P. R., Holden, H., and Andrews, P. W. (2007) Adaptation to culture of human embryonic stem cells and oncogenesis in vivo, *Nat Biotechnol 25*, 207–

215.

72. Närvä, E., Autio, R., Rahkonen, N., Kong, L., Harrison, N., Kitsberg, D., Borghese, L., Itskovitz-Eldor, J., Rasool, O., Dvorak, P., Hovatta, O., Otonkoski, T., Tuuri, T., Cui, W., Brüstle, O., Baker, D., Maltby, E., Moore, H. D., Benvenisty, N., Andrews, P. W., Yli-Harja, O., and Lahesmaa, R. (2010) High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity, *Nat Biotechnol 28*, 371–377.

73. Pasi, C. E., Dereli-Öz, A., Negrini, S., Friedli, M., Fragola, G., Lombardo, A., Van Houwe, G., Naldini, L., Casola, S., Testa, G., Trono, D., Pelicci, P. G., and Halazonetis, T. D. (2011) Genomic instability in induced stem cells, *Cell Death Differ. 18*, 745–753.

74. Zhang, D., Jiang, W., Liu, M., Sui, X., Yin, X., Chen, S., Shi, Y., and Deng, H. (2009) Highly efficient differentiation of human ES cells and iPS cells into mature pancreatic insulin-producing cells, *Cell Res 19*, 429–438.

75. Taura, D., Noguchi, M., Sone, M., Hosoda, K., Mori, E., Okada, Y., Takahashi, K., Homma, K., Oyamada, N., Inuzuka, M., Sonoyama, T., Ebihara, K., Tamura, N., Itoh, H., Suemori, H., Nakatsuji, N., Okano, H., Yamanaka, S., and Nakao, K. (2009) Adipogenic differentiation of human induced pluripotent stem cells: comparison with that of human embryonic stem cells, *FEBS Letters 583*, 1029–1033.

76. Dimos, J. T., Rodolfa, K. T., Niakan, K. K., Weisenthal, L. M., Mitsumoto, H., Chung, W., Croft, G. F., Saphier, G., Leibel, R., Goland, R., Wichterle, H.,

Henderson, C. E., and Eggan, K. (2008) Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons, *Science 321*, 1218–1221.

77. Li, Y., Tsai, Y.-T., Hsu, C.-W., Erol, D., Yang, J., Wu, W.-H., Davis, R. J., Egli, D., and Tsang, S. H. (2012) Long-term safety and efficacy of human-induced pluripotent stem cell (iPS) grafts in a preclinical model of retinitis pigmentosa, *Mol. Med. 18*, 1312–1319.

78. Wang, D., Haviland, D. L., Burns, A. R., Zsigmond, E., and Wetsel, R. A. (2007) A pure population of lung alveolar epithelial type II cells derived from human embryonic stem cells, *Proc. Natl. Acad. Sci. U.S.A. 104*, 4449–4454.

79. Longmire, T. A., Ikonomou, L., Hawkins, F., Christodoulou, C., Cao, Y., Jean, J. C., Kwok, L. W., Mou, H., Rajagopal, J., Shen, S. S., Dowton, A. A., Serra, M., Weiss, D. J., Green, M. D., Snoeck, H.-W., Ramirez, M. I., and Kotton, D. N. (2012) Efficient derivation of purified lung and thyroid progenitors from embryonic stem cells, *Cell Stem Cell 10*, 398–411.

80. Mou, H., Zhao, R., Sherwood, R., Ahfeldt, T., Lapey, A., Wain, J., Sicilian, L., Izvolsky, K., Musunuru, K., Cowan, C., and Rajagopal, J. (2012) Generation of multipotent lung and airway progenitors from mouse ESCs and patient-specific cystic fibrosis iPSCs, *Cell Stem Cell 10*, 385–397.

81. Wong, A. P., Bear, C. E., Chin, S., Pasceri, P., Thompson, T. O., Huan, L.-J., Ratjen, F., Ellis, J., and Rossant, J. (2012) Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein, *Nat Biotechnol 30*, 876–882.

82. Hacein-Bey-Abina, S., Garrigue, A., Wang, G. P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., Asnafi, V., MacIntyre, E., Dal Cortivo, L., Radford, I., Brousse, N., Sigaux, F., Moshous, D., Hauer, J., Borkhardt, A., Belohradsky, B. H., Wintergerst, U., Velez, M. C., Leiva, L., Sorensen, R., Wulffraat, N., Blanche, S., Bushman, F. D., Fischer, A., and Cavazzana-Calvo, M. (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1, *J. Clin. Invest. 118*, 3132–3142.

83. Stein, S., Ott, M. G., Schultze-Strasser, S., Jauch, A., Burwinkel, B., Kinner, A., Schmidt, M., Krämer, A., Schwäble, J., Glimm, H., Koehl, U., Preiss, C., Ball, C., Martin, H., Göhring, G., Schwarzwaelder, K., Hofmann, W.-K., Karakaya, K., Tchatchou, S., Yang, R., Reinecke, P., Kühlcke, K., Schlegelberger, B., Thrasher, A. J., Hoelzer, D., Seger, R., Kalle, von, C., and Grez, M. (2010) Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease, *Nat Med 16*, 198–204.

84. Lloyd, A., Plaisier, C. L., Carroll, D., and Drews, G. N. (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis, *Proc. Natl. Acad. Sci. U.S.A. 102*, 2232–2237.

85. Jasin, M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases, *Trends Genet. 12*, 224–228.

86. Porteus, M. H., and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells, *Science 300*, 763.

87. Rouet, P., Smih, F., and Jasin, M. (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells, *Proc.*

Natl. Acad. Sci. U.S.A. 91, 6064–6068.

88. Kim, Y. G., Cha, J., and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain, *Proc. Natl. Acad. Sci. U.S.A. 93*, 1156–1160.

89. Mandell, J. G., and Barbas, C. F. (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases, *Nucleic Acids Research 34*, W516–23.

90. Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G., and Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases, *Molecular and Cellular Biology 21*, 289–297.

91. Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y.-L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A., Gregory, P. D., Pabo, C. O., and Rebar, E. J. (2007) An improved zinc-finger nuclease architecture for highly specific genome editing, *Nat Biotechnol 25*, 778–785.

92. Pruett-Miller, S. M., Connelly, J. P., Maeder, M. L., Joung, J. K., and Porteus, M. H. (2008) Comparison of zinc finger nucleases for use in gene targeting in mammalian cells, *Mol Ther 16*, 707–717.

93. Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V., Lee, Y.-L., Guschin, D. Y., Rupniewski, I., Waite, A. J., Carpenito, C., Carroll, R. G., Orange, J. S., Urnov, F. D., Rebar, E. J., Ando, D., Gregory, P. D., Riley, J. L., Holmes, M. C., and June, C. H. (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using

zinc-finger nucleases, *Nat Biotechnol 26*, 808–816.

94. Alwin, S., GERE, M., Guhl, E., Effertz, K., Barbas, C., Segal, D., Weitzman, M., and Cathomen, T. (2005) Custom Zinc-Finger Nucleases for Use in Human Cells, *Mol Ther 12*, 610–617.

95. Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J. K., and Carroll, D. (2006) Efficient gene targeting in Drosophila with zinc-finger nucleases, *Genetics 172*, 2391–2403.

96. Cornu, T. I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichtinger, M., Joung, J. K., Joung, J. K., and Cathomen, T. (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases, *Mol Ther 16*, 352–358.

97. Carroll, D. (2008) Progress and prospects: zinc-finger nucleases as gene therapy agents, *Gene Ther. 15*, 1463–1468.

98. Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R. C., Katibah, G. E., Amora, R., Boydston, E. A., Zeitler, B., Meng, X., Miller, J. C., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D., and Jaenisch, R. (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases, *Nat Biotechnol 27*, 851–857.

99. Sander, J. D., Dahlborg, E. J., Goodwin, M. J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S. J., Blackburn, J. S., Thibodeau-Beganny, S., Qi, Y., Pierick, C. J., Hoffman, E., Maeder, M. L., Khayter, C., Reyon, D., Dobbs, D., Langenau, D. M., Stupar, R. M., Giraldez, A. J., Voytas, D. F., Peterson, R. T., Yeh, J.-R. J., and Joung, J. K. (2011) Selection-free zinc-finger-nuclease engineering by

context-dependent assembly (CoDA), *Nat Meth 8*, 67–69.

100. Gupta, A., Meng, X., Zhu, L. J., Lawson, N. D., and Wolfe, S. A. (2011) Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases, *Nucleic Acids Research 39*, 381–392.

101. Pattanayak, V., Ramirez, C. L., Joung, J. K., and Liu, D. R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection, *Nat Meth 8*, 765–770.

102. Szczepek, M., Brondani, V., Büchel, J., Serrano, L., Segal, D. J., and Cathomen, T. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases, *Nat Biotechnol 25*, 786–793.

103. Urnov, F. D., Miller, J. C., Lee, Y.-L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., and Holmes, M. C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases, *Nature 435*, 646–651.

104. DeKelver, R. C., Choi, V. M., Moehle, E. A., Paschon, D. E., Hockemeyer, D., Meijsing, S. H., Sancak, Y., Cui, X., Steine, E. J., Miller, J. C., Tam, P., Bartsevich, V. V., Meng, X., Rupniewski, I., Gopalan, S. M., Sun, H. C., Pitz, K. J., Rock, J. M., Zhang, L., Davis, G. D., Rebar, E. J., Cheeseman, I. M., Yamamoto, K. R., Sabatini, D. M., Jaenisch, R., Gregory, P. D., and Urnov, F. D. (2010) Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome, *Genome Research 20*, 1133–1142.

105. Lombardo, A., Cesana, D., Genovese, P., Di Stefano, B., Provasi, E., Colombo,

D. F., Neri, M., Magnani, Z., Cantore, A., Riso, Lo, P., Damo, M., Pello, O. M., Holmes, M. C., Gregory, P. D., Gritti, A., Broccoli, V., Bonini, C., and Naldini, L. (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer, *Nat Meth 8*, 861–869.

106. Moehle, E. A., Moehle, E. A., Rock, J. M., Rock, J. M., Lee, Y.-L., Lee, Y. L., Jouvenot, Y., Jouvenot, Y., DeKelver, R. C., DeKelver, R. C., Gregory, P. D., Gregory, P. D., Urnov, F. D., Urnov, F. D., Holmes, M. C., and Holmes, M. C. (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases, *Proc. Natl. Acad. Sci. U.S.A. 104*, 3055–3060.

107. Deng, C., and Capecchi, M. R. (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus, *Molecular and Cellular Biology 12*, 3365–3371.

108. Thomas, K. R., and Capecchi, M. R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells, *Cell*, Elsevier *51*, 503–512.

109. Waldman, A. S., and Liskay, R. M. (1988) Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology, *Molecular and Cellular Biology 8*, 5350–5357.

110. Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A., and Jasin, M. (1998) Gene conversion tracts from double-strand break repair in mammalian cells, *Molecular and Cellular Biology 18*, 93–101.

111. Garate, Z., Davis, B. R., Quintana-Bustamante, O., and Segovia, J. C. (2013) New frontier in regenerative medicine: site-specific gene correction in patient-

specific induced pluripotent stem cells, *Human Gene Therapy 24*, 571–583.

112. Lombardo, A., Genovese, P., Beausejour, C. M., Colleoni, S., Lee, Y.-L., Kim, K. A., Ando, D., Urnov, F. D., Galli, C., Gregory, P. D., Holmes, M. C., and Naldini, L. (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery, *Nat Biotechnol 25*, 1298–1306.

113. Yusa, K., Rashid, S. T., Strick-Marchand, H., Varela, I., Liu, P.-Q., Paschon, D. E., Miranda, E., Ordóñez, A., Hannan, N. R. F., Rouhani, F. J., Darche, S., Alexander, G., Marciniak, S. J., Fusaki, N., Hasegawa, M., Holmes, M. C., Di Santo, J. P., Lomas, D. A., Bradley, A., and Vallier, L. (2011) Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells, *Nature 478*, 391–394.

114. Zou, J., Mali, P., Huang, X., Dowey, S. N., and Cheng, L. (2011) Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease, *Blood 118*, 4599–4608.

115. Chang, C.-J., and Bouhassira, E. E. (2012) Zinc-finger nuclease-mediated correction of α-thalassemia in iPS cells, *Blood 120*, 3906–3914.

116. Hockemeyer, D., Wang, H., Kiani, S., Lai, C. S., Gao, Q., Cassady, J. P., Cost, G. J., Zhang, L., Santiago, Y., Miller, J. C., Zeitler, B., Cherone, J. M., Meng, X., Hinkley, S. J., Rebar, E. J., Gregory, P. D., Urnov, F. D., and Jaenisch, R. (2011) Genetic engineering of human pluripotent cells using TALE nucleases, *Nat Biotechnol 29*, 731–734.

117. Ding, Q., Lee, Y.-K., Schaefer, E. A. K., Peters, D. T., Veres, A., Kim, K., Kuperwasser, N., Motola, D. L., Meissner, T. B., Hendriks, W. T., Trevisan, M.,

Gupta, R. M., Moisan, A., Banks, E., Friesen, M., Schinzel, R. T., Xia, F., Tang, A., Xia, Y., Figueroa, E., Wann, A., Ahfeldt, T., Daheron, L., Zhang, F., Rubin, L. L., Peng, L. F., Chung, R. T., Musunuru, K., and Cowan, C. A. (2013) A TALEN genome-editing system for generating human stem cell-based disease models, *Cell Stem Cell 12*, 238–251.

118. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013) RNA-guided human genome engineering via Cas9, *Science 339*, 823–826.

119. Trapnell, B. C., Chu, C. S., Paakko, P. K., Banks, T. C., Yoshimura, K., Ferrans, V. J., Chernick, M. S., and Crystal, R. G. (1991) Expression of the cystic fibrosis transmembrane conductance regulator gene in the respiratory tract of normal individuals and individuals with cystic fibrosis, *Proc. Natl. Acad. Sci. U.S.A. 88*, 6565–6569.

120. Park, I.-H., and Daley, G. Q. (2009) Human iPS cell derivation/reprogramming, *Curr Protoc Stem Cell Biol Chapter 4*, Unit 4A.1.

121. Stewart, S. A., Dykxhoorn, D. M., Palliser, D., Mizuno, H., Yu, E. Y., An, D. S., Sabatini, D. M., Chen, I. S. Y., Hahn, W. C., Sharp, P. A., Weinberg, R. A., and Novina, C. D. (2003) Lentivirus-delivered stable gene silencing by RNAi in primary cells, *RNA 9*, 493–501.

122. Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., Ko, M. S. H., and Niwa, H. (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells, *Nat Cell Biol 9*, 625–635.

123. Le, Y., Miller, J. L., and Sauer, B. (1999) GFPcre fusion vectors with enhanced expression, *Anal. Biochem. 270*, 334–336.

124. Matsuda, T., and Cepko, C. L. (2007) Controlled expression of transgenes introduced by in vivo electroporation, *Proc. Natl. Acad. Sci. U.S.A. 104*, 1027–1032.

125. D'Amour, K. A., Agulnick, A. D., Eliazer, S., Kelly, O. G., Kroon, E., and Baetge, E. E. (2005) Efficient differentiation of human embryonic stem cells to definitive endoderm, *Nat Biotechnol 23*, 1534–1541.

126. Van Haute, L., De Block, G., Liebaers, I., Sermon, K., and De Rycke, M. (2009) Generation of lung epithelial-like tissue from human embryonic stem cells, *Respir Res 10*, 105.

127. Green, M. D., Chen, A., Nostro, M.-C., d'Souza, S. L., Schaniel, C., Lemischka, I. R., Gouon-Evans, V., Keller, G., and Snoeck, H.-W. (2011) Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells, *Nat Biotechnol 29*, 267–272.

128. Gouon-Evans, V., Boussemart, L., Gadue, P., Nierhoff, D., Koehler, C. I., Kubo, A., Shafritz, D. A., and Keller, G. (2006) BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm, *Nat Biotechnol 24*, 1402–1411.

129. Cradick, T. J., Ambrosini, G., Iseli, C., Bucher, P., and McCaffrey, A. P. (2011) ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites, *BMC Bioinformatics 12*, 152.

130. Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z. D., Ziller,

M., Croft, G. F., Amoroso, M. W., Oakley, D. H., Gnirke, A., Eggan, K., and Meissner, A. (2011) Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines, *Cell 144*, 439–452.

131. Kubo, A., Shinozaki, K., Shannon, J. M., Kouskoff, V., Kennedy, M., Woo, S., Fehling, H. J., and Keller, G. (2004) Development of definitive endoderm from embryonic stem cells in culture, *Development 131*, 1651–1662.

132. Coraux, C., Nawrocki-Raby, B., Hinnrasky, J., Kileztky, C., Gaillard, D., Dani, C., and Puchelle, E. (2005) Embryonic stem cells generate airway epithelial tissue, *Am. J. Respir. Cell Mol. Biol. 32*, 87–92.

133. Perl, A.-K. T., Kist, R., Shan, Z., Scherer, G., and Whitsett, J. A. (2005) Normal lung development and function after Sox9 inactivation in the respiratory epithelium, *Genesis 41*, 23–32.

134. Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., and Thomson, J. A. (2007) Induced pluripotent stem cell lines derived from human somatic cells, *Science 318*, 1917–1920.

135. Hall, B., Limaye, A., and Kulkarni, A. B. (2009) Overview: generation of gene knockout mice, *Curr Protoc Cell Biol Chapter 19*, Unit 19.12 19.12.1–17.

136. Lee, C. M., Flynn, R., Hollywood, J. A., Scallan, M. F., and Harrison, P. T. (2012) Correction of the ΔF508 Mutation in the Cystic Fibrosis Transmembrane Conductance Regulator Gene by Zinc-Finger Nuclease Homology-Directed Repair, *Biores Open Access 1*, 99–108.

137. Ramalingam, S., London, V., Kandavelou, K., Cebotaru, L., Guggino, W., Civin,

C., and Chandrasegaran, S. (2013) Generation and genetic engineering of human induced pluripotent stem cells using designed zinc finger nucleases, *Stem Cells Dev. 22*, 595–610.

138. Verkman, A. S., and Jayaraman, S. (2002) Fluorescent indicator methods to assay functional CFTR expression in cells, *Methods Mol. Med. 70*, 187–196.

139. Munkonge, F., Alton, E. W. F. W., Andersson, C., Davidson, H., Dragomir, A., Edelman, A., Farley, R., Hjelte, L., McLachlan, G., Stern, M., and Roomans, G. M. (2004) Measurement of halide efflux from cultured and primary airway epithelial cells using fluorescence indicators, *J. Cyst. Fibros. 3*, 171–176.

140. Brown, C. R., Hong-Brown, L. Q., Biwersi, J., Verkman, A. S., and Welch, W. J. (1996) Chemical chaperones correct the mutant phenotype of the delta F508 cystic fibrosis transmembrane conductance regulator protein, *Cell Stress Chaperones 1*, 117–125.

141. Price, A. P., England, K. A., Matson, A. M., Blazar, B. R., and Panoskaltsis-Mortari, A. (2010) Development of a decellularized lung bioreactor system for bioengineering the lung: the matrix reloaded, *Tissue Eng Part A 16*, 2581–2591.

142. Ott, H. C., Clippinger, B., Conrad, C., Schuetz, C., Pomerantseva, I., Ikonomou, L., Kotton, D., and Vacanti, J. P. (2010) Regeneration and orthotopic transplantation of a bioartificial lung, *Nat Med 16*, 927–933.

143. Daly, A. B., Wallis, J. M., Borg, Z. D., Bonvillain, R. W., Deng, B., Ballif, B. A., Jaworski, D. M., Allen, G. B., and Weiss, D. J. (2012) Initial binding and recellularization of decellularized mouse lung scaffolds with bone marrow-derived mesenchymal stromal cells, *Tissue Eng Part A 18*, 1–16.

144. Petersen, T. H., Calle, E. A., Zhao, L., Lee, E. J., Gui, L., Raredon, M. B., Gavrilov, K., Yi, T., Zhuang, Z. W., Breuer, C., Herzog, E., and Niklason, L. E. (2010) Tissue-engineered lungs for in vivo implantation, *Science 329*, 538–541.

145. Cyranoski, D. (2013) Japan to start stem-cell study on humans, *Nature News Blog*, http://blogs.nature.com/news/2013/07/japan-to-start-stem-cell-study-on-humans.html.

Vita

Jacquelin (Jacky) Hanh Bui was born in Houston, Tx on May 24, 1985 to Le Roberts and Lang Bui. Jacky was raised in Houston and went to Debakey High School for Health Professions. She then attended Trinity University in San Antonio, Tx, from which she obtained a B.S. in Biology in May of 2007. The following fall she matriculated into the University of Texas Health Science Center Graduate School for Biomedical Sciences where she joined the laboratory of Dr. Brian Davis in 2008. She has since then been working on her dissertation: implementing two rather novel technologies towards an alternative gene and cellular therapy for Cystic Fibrosis.